

Health Technology Assessment Program: Selected Technologies 2014

List of Contents

1. Director's selection letter
2. Topic selection background information
3. Public comments on proposed topics and responses



STATE OF WASHINGTON
HEALTH CARE AUTHORITY

626 8th Avenue, SE • P.O. Box 45502 • Olympia, Washington 98504-5502

April 4, 2014

To whom it may concern:

SUBJECT: Health Technology Assessment Topic Selection 2014

As the Director of the Washington State Health Care Authority (HCA) and per the Health Technology Assessment law (70.14 RCW), I select technologies for review by the program in consultation with other agencies and the Health Technology Clinical Committee. Technologies are selected when there are concerns about safety, efficacy or value (cost-effectiveness), when state expenditures are or could be high, and there is adequate evidence to conduct a review. Technologies are selected for re-review when new evidence is available that could change a previous determination. In addition, anyone may petition for a technology review.

For the current selection cycle, I have reviewed the proposed topics as well as the comments received from the interested individuals and groups who responded in the first comment period (February 28 through March 17, 2014). Based on the information provided by the Health Technology Assessment (HTA) Program, and the recommendations from staff in HCA, Department of Labor and Industries and Department of Corrections, I have selected the following technologies for review:

| | Topic | Primary Criteria ¹ | | |
|---|--|-------------------------------|----------|-------------|
| | | Safety | Efficacy | Cost |
| 1 | Testosterone Testing Policy Context/Reason for selection: High diffusion, concern for: limited evidence for benefit and risk levels, variable and inaccurate test results with unclear normal levels, high estimated treatment costs. | High | High | Medium/High |
| 2 | Bariatric Surgery for Overweight/ Obese Policy Context/Reason for selection: Centers for Medicare and Medicaid Services (CMS) National Coverage Decision (NCD) is limited to morbid obesity with indications expanding into overweight categories. | High | Medium | High |
| 3 | Imaging for Rhinosinusitis Policy Context/Reason for selection: Concerns: usage expansion in low risk, high prevalence conditions increases radiation exposure, special consideration for children. | Medium | High | Medium |
| 4 | Appropriate Breast Imaging for Breast Cancer Screening in Special Conditions Policy Context/Reason for selection: Concerns include: new concerns for efficacy in special populations (e.g., high risk, dense tissue), high state usage. | Low | High | Medium/High |
| 5 | Tympanostomy Tubes Policy Context/Reason for selection: Concerns include: variation in use of treatment across different populations (e.g., socioeconomic status), potential overuse in some populations, high state usage. | Medium | High | Medium/Low |

¹ [Link to prioritization criteria.](#)

April 4, 2014
Page 2

Additionally, I have selected Lumbar Spinal Fusion for re-review based on the newly available published evidence.

Upon publication of the selected list of technologies, a 30-day comment period will begin whereby any interested person or group may provide information relevant to review of these topics. HTA will begin work to review these technologies following this comment period.

Should you have any questions or concerns, please contact Josh Morse, HTA Program manager by telephone at 360-725-0839 or via email at Josiah.morse@hca.wa.gov.

Sincerely,

A black rectangular redaction box covering the signature of Dorothy F. Teetet.

Dorothy F. Teetet, MHA
Director

Selected Technologies

| Topic | | Primary Criteria ¹ | | |
|-------|--|-------------------------------|----------|-------------|
| | | Safety | Efficacy | Cost |
| 1 | Testosterone Testing Policy Context/Reason for selection: High diffusion, concern for: limited evidence for benefit and risk levels, variable and inaccurate test results with unclear normal levels, high estimated treatment costs. | High | High | Medium/High |
| 2 | Bariatric Surgery for Overweight/ Obese Policy Context/Reason for selection: Center for Medicare & Medicaid Services (CMS) National Coverage Decision (NCD) is limited to morbid obesity with indications expanding into overweight categories. | High | Medium | High |
| 3 | Imaging for Rhinosinusitis Policy Context/Reason for selection: Concerns: usage expansion in low risk, high prevalence conditions increases radiation exposure, special consideration for children. | Medium | High | Medium |
| 4 | Appropriate Breast Imaging for Breast Cancer Screening in Special Conditions Policy Context/Reason for selection: Concerns include: new concerns for efficacy in special populations (e.g., high risk, dense tissue), high state usage. | Low | High | Medium/High |
| 5 | Tympanostomy Tubes Policy Context/Reason for selection: Concerns include: variation in use of treatment across different populations (e.g., socioeconomic status), potential overuse in some populations, high state usage. | Medium | High | Medium/Low |

¹ [Link to prioritization criteria.](#)

Technologies Considered, Not Proposed

| Topic | |
|-------|--|
| 1 | CT for general abdominal and pelvic pain in adults and children Concern for special populations. Data indicates low utilization. |
| 2 | Radiologic screening for carotid artery stenting in asymptomatic adults U.S. Preventive Services Task Force rating of 'D'. Concern for potential overuse of procedure. Very low usage. |
| 3 | Insulin pumps for type I or type II diabetes CMS NCD for external pumps. No NCD for internal pumps. Low utilization. |

Technologies Considered, Not Proposed

| Topic | |
|----------|---|
| 4 | Renal stenting Concerns about effectiveness. Possible future review. |
| 5 | High frequency chest wall compression devices High risk population(s). Low utilization. High-cost device. Limited evidence for some conditions. |
| 6 | Specialty eye imaging for patients without symptoms Topic needs further definition through review of guidelines. |
| 7 | Biologics for rheumatoid arthritis Further definition required. |
| 8 | Platelet rich plasma injections Currently limited use, mixed evidence results. |
| 9 | Hepatocellular carcinoma screening in chronic viral hepatitis Studies indicate efficacy may be adequate. |

Re-Review Technologies:

Technologies are considered for re-review at least once every eighteen months based on availability of new evidence that may change the decision. (*Detailed criteria are included below*). All technologies with determinations beyond 18 months since the final determination previously reviewed by the health technology clinical committee (HTCC) are listed below, along with information on whether they have been selected for re-review.

| Topic | | Originally Reviewed | Recommended for Re-review |
|----------|--|---------------------|---------------------------|
| 1 | Pediatric Bariatric Surgery Surveillance conducted. Original evidence conclusions similar to more recent findings. | August 2007 | No |
| 2 | Artificial Discs New literature, new indications beyond scope of original review for 2 level cervical disc replacement. | October 2008 | No |
| 3 | Spinal Fusion New literature identified that could change evidence findings/ original determination (bibliography attached). | November 2007 | Yes |
| 4 | Spinal Cord Stimulators Surveillance conducted. New studies identified are not likely to change original determination. | August 2010 | No |
| 5 | Vagal Nerve Stimulation Surveillance conducted. | October 2009 | No |

For the current period, the program has not received or identified new evidence to support review of the following:

| | Topic | Date of Last Search or Re-Review |
|-----------|--|---|
| 1 | Virtual Colonoscopy or Computed Tomographic Colonography | February 2008 |
| 2 | Discography | August 2008 |
| 3 | Arthroscopic Knee Surgery | October 2008 |
| 4 | Implantable Infusion Pumps | October 2008 |
| 5 | Computed Tomographic Angiography | May 2009 |
| 6 | Cardiac Stents | August 2009 |
| 7 | Bone Growth Stimulators | October 2009 |
| 8 | Electrical Neural Stimulation | November 2009 |
| 9 | Calcium Scoring | May 2010 |
| 10 | Breast MRI | October 2010 |
| 11 | Knee Joint Replacement or Knee Arthroplasty | December 2010 |
| 12 | Routine Ultrasound for Pregnancy | December 2010 |
| 13 | Vertebroplasty, Kyphoplasty, Sacroplasty | March 2011 |
| 14 | Spinal Injections-Therapeutic | June 2011 |
| 15 | Glucose Monitoring | June 2011 |
| 16 | Applied Behavioral Analysis Therapy for Autism | September 2011 |
| 17 | Femoroacetabular Impingement Syndrome | November 2011 |
| 18 | Positron Emission Tomography Scans for Lymphoma | November 2011 |
| 19 | Microprocessor-controlled Lower Limb Prosthetics | March 2012 |
| 20 | Osteochondral Allograft / Autograft Transplantation | March 2012 |
| 21 | Sleep Apnea Diagnosis and Treatment | May 2012 |
| 22 | Bone Morphogenetic Protein | May 2012 |

Agency Medical Director's Brief on Re-review of lumbar fusion for uncomplicated degenerative disc disease

The **prior HTA assessment of lumbar fusion** was completed 11/16/07, with final determination updated on 2/15/2008. The final literature search for that HTA review was conducted in August, 2007. Only one included study was published in 2007 (Martin et al, 2007), all other cited studies were published prior to 2007

The studies included for KQ1 (does pain and function improve?) only included randomized trials, but no high quality or population-based observational studies for effectiveness were included. For KQ3, on special populations, the vendor report excluded all existing studies on lumbar fusion in workers compensation, so the specific question related to that special population could not be answered.

Hayes conducted a review of studies published July, 2007 through Nov 6, 2012.

This review may have missed studies published earlier in 2007. The Hayes report points to an Agency for Healthcare Research and Quality report from 2011, and includes 4 more systematic reviews of efficacy or effectiveness. The overall Hayes impression was a somewhat more positive view of lumbar fusion; Hayes did not conduct any in depth new assessment of safety or cost-effectiveness.

We conducted an updated review searching for efficacy and effectiveness studies published since the Hayes review end date of Nov 2012. We used key words: lumbar fusion; lumbar arthrodesis; outcomes; meta-analysis; randomized. Several additional systematic reviews and meta-analyses of high quality RCTs pertinent to efficacy and effectiveness, one new prospective observational study (Mirza et al, 2013), and a longer term observational study of outcome of 3 of the RCTs included in original HTA (Mannion et al) have been published since the Hayes review:

J Spinal Disord Tech. 2013 Dec 15. [Epub ahead of print]

Lumbar Fusion versus Non-operative Management for Treatment of Discogenic Low Back Pain: A Systematic Review and Meta-analysis of Randomized Controlled Trials.

Bydon M, De la Garza-Ramos R, Macki M, Baker A, Gokaslan ZL, Bydon A.

Author information

Abstract

STUDY DESIGN::

Systematic review and meta-analysis of randomized controlled trials (RCTs).

OBJECTIVE::

To evaluate the current evidence comparing lumbar fusion to non-operative

management for the treatment of chronic discogenic low back pain.

BACKGROUND AND CONTEXT::

Discogenic low back pain is a common and sometimes disabling condition. When the condition becomes chronic and intractable, spinal fusion may play a role.

METHODS::

A systematic review of the literature was conducted using the PubMed and CENTRAL databases. We included RCTs that compared lumbar fusion to non-operative management for the treatment of adult patients with chronic discogenic low back pain. A meta-analysis was conducted to assess the improvement in back pain based on the ODI.

RESULTS::

Five RCTs met our inclusion criteria. A total of 707 patients were divided into lumbar fusion (n=523) and conservative management (n=134). Although inclusion/exclusion criteria were relatively similar across studies, surgical techniques and conservative management protocols varied. The pooled mean difference in ODI (final ODI minus initial ODI) between the non-operative and lumbar fusion groups across all studies was -7.39 points [95% CI: -20.26, 5.47] in favor of lumbar fusion, but this difference was not statistically significant (P=0.26).

CONCLUSIONS::

Despite the significant improvement in ODI in the lumbar fusion groups in three studies, pooled data revealed no significant difference when compared to the non-operative group. Although there was an overall improvement of 7.39 points in the ODI in favor of lumbar fusion, it is unclear that this change in ODI would lead to a clinically significant difference. Prospective randomized trials comparing a specific surgical technique versus a structured physical therapy program may improve evidence quality. Until then, either operative intervention via lumbar fusion or non-operative management and physical therapy remain two acceptable treatment methods for intractable low back pain.

Spine J. 2013 Nov;13(11):1438-48. doi: 10.1016/j.spinee.2013.06.101. Epub 2013 Nov 5.

Comparison of spinal fusion and nonoperative treatment in patients with chronic low back pain: long-term follow-up of three randomized controlled trials.

Mannion AF, Brox JL, Fairbank JC.

Author information

Abstract

BACKGROUND CONTEXT:

Chronic low back pain (cLBP) represents a major challenge to our health care systems. The relative efficacy of surgery over nonoperative treatment for the treatment of cLBP remains controversial, and little is known of the long-term

comparative outcomes.

PURPOSE:

To compare the clinical outcome at long-term follow-up (LTFU) of patients who were randomized with either spinal fusion or multidisciplinary cognitive-behavioral and exercise rehabilitation for cLBP.

STUDY DESIGN/SETTING:

Long-term clinical follow-up of three multicenter randomized controlled trials (RCTs) of surgery (instrumented or noninstrumented fusion, stabilization) versus nonoperative treatment (multidisciplinary cognitive-behavioral and exercise rehabilitation) in Norway and the United Kingdom.

PATIENT SAMPLE:

A total of 473 patients with cLBP of at least 1 year's duration who were all considered candidates for spinal fusion.

OUTCOME MEASURES:

The primary outcome was the Oswestry Disability Index (ODIv2.1a for the United Kingdom and ODIv1 for Norway) score measured at LTFU. Secondary outcomes included visual analog scale (VAS) pain intensity, pain frequency, pain medication use, work status, EuroQol VAS for health-related quality of life, satisfaction with care, and global treatment outcome at LTFU.

METHODS:

Patients who consented to LTFU (average 11.4 [range 8-15] years after the initial treatment) completed the outcome questionnaires.

RESULTS:

Of 473 enrolled patients, 261 (55%) completed LTFU, 140/242 patients randomized to receive surgery and 121/231 randomized to receive multidisciplinary cognitive-behavioral and exercise rehabilitation. The intention-to-treat analysis showed no statistically or clinically significant differences between treatment groups for ODI scores at LTFU (adjusted for baseline ODI, previous surgery, duration of LBP, sex, age, and smoking habit): the mean adjusted treatment effect of fusion was -0.7 points on the 0-100 ODI scale (95% confidence interval [CI], -5.5 to 4.2). An as-treated analysis similarly demonstrated no advantage of surgery (treatment effect, -0.8 points on the ODI (95% CI, -5.9 to 4.3). The results for the secondary outcomes were largely consistent with those of the ODI, showing no relevant group differences.

CONCLUSIONS:

After an average of 11 years follow-up, there was no difference in patient self-rated outcomes between fusion and multidisciplinary cognitive-behavioral and exercise rehabilitation for cLBP. The results suggest that, given the increased risks of surgery and the lack of deterioration in nonoperative outcomes over time, the use of lumbar fusion in cLBP patients should not be favored in health care systems where multidisciplinary cognitive-behavioral and exercise rehabilitation programmes are available.

Int J Rehabil Res. 2013 Jun 29. [Epub ahead of print]

Lumbar fusion compared with conservative treatment in patients with chronic low back pain: a meta-analysis.

Saltychev M, Eskola M, Laimi K.

Author information

Abstract

We assess the effect of lumbar fusion (LF) in reducing disability among patients with chronic low back pain (CLBP) compared with conservative treatment and to weigh the clinical significance of this effect. We conducted a random-effect meta-analysis on the basis of a systematic review with research quality grading according to Grading of Recommendations Assessment, Development and Evaluation (GRADE). The studies included were retrieved from MEDLINE and Cochrane CENTRAL databases from 1990 till January 2013. Randomized or nonrandomized controlled studies were included if the study participants had a history of CLBP because of degenerative spinal diseases and had been treated with LF. A study was included if it compared LF with conservative treatment. The outcome measure was a change in the Oswestry Disability Index (ODI) score during a follow-up. The meta-analysis included data on 666 patients (402 cases) who participated in four randomized-controlled trials. The ODI score reduced in the LF and the control groups. The mean reduction in the ODI score in the follow-up of 1.5 years was -2.91 (95% confidence interval -6.66 to 0.84) in favor of LF. The difference between groups was statistically and clinically insignificant. Test for heterogeneity indicated that study imputation would favor LF but the imputed result would still be clinically insignificant with an estimated corrected reduction of ODI score of -5.51 (95% confidence interval -5.78 to -5.24). There is strong evidence that LF is not more effective than conservative treatment in reducing perceived disability because of CLBP among patients with degenerative spinal diseases. It is unlikely that further research on the subject would considerably affect this conclusion.

Spine J. 2013 Nov;13(11):1421-33. doi: 10.1016/j.spinee.2013.05.047. Epub 2013 Jul 23.

One-year outcomes of surgical versus nonsurgical treatments for discogenic back pain: a community-based prospective cohort study.

Mirza SK, Deyo RA, Heagerty PJ, Turner JA, Martin BI, Comstock BA.

Author information

Abstract

BACKGROUND CONTEXT:

The clinical entity "discogenic back pain" remains controversial at fundamental

levels, including its pathophysiology, diagnostic criteria, and optimal treatment. This is true despite availability of four randomized trials comparing the efficacy of surgical and nonsurgical treatments. One trial showed benefit for lumbar fusion compared with unstructured nonoperative care, and three others showed roughly similar results for lumbar surgery and structured rehabilitation.

PURPOSE:

To compare outcomes of community-based surgical and nonsurgical treatments for patients with chronic back pain attributed to degeneration at one or two lumbar disc levels.

DESIGN:

Prospective observational cohort study.

PATIENT SAMPLE:

Patients presenting with axial back pain to academic and private practice orthopedic surgeons and neurosurgeons in a large metropolitan area.

OUTCOME MEASURES:

Roland-Morris back disability score (primary outcome), current rating of overall pain severity on a numerical scale, back and leg pain bothersomeness measures, the physical function scale of the short-form 36 version 2 questionnaire, use of medications for pain, work status, emergency department visits, hospitalizations, and further surgery.

METHODS:

Patients receiving spine surgery within 6 months of enrollment were designated as the "surgical treatment" group and the remainder as "nonsurgical treatment." Outcomes were assessed at 3, 6, 9, and 12 months after enrollment.

RESULTS:

We enrolled 495 patients with discogenic back pain presenting for initial surgical consultation in offices of 16 surgeons. Eighty-six patients (17%) had surgery within 6 months of enrollment. Surgery consisted of instrumented fusion (79%), disc replacement (12%), laminectomy, or discectomy (9%). Surgical patients reported more severe pain and physical disability at baseline and were more likely to have had prior surgery. Adjusting for baseline differences among groups, surgery showed a limited benefit over nonsurgical treatment of 5.4 points on the modified (23-point) Roland disability questionnaire (primary outcome) 1 year after enrollment. Using a composite definition of success incorporating 30% improvement in the Roland score, 30% improvement in pain, no opioid pain medication use, and working (if relevant), the 1-year success rate was 33% for surgery and 15% for nonsurgical treatment. The rate of reoperation was 11% in the surgical group; the rate of surgery after treatment designation in the nonsurgical group was 6% at 12 months after enrollment.

CONCLUSIONS:

The surgical group showed greater improvement at 1 year compared with the nonsurgical group, although the composite success rate for both treatment groups was only fair. The results should be interpreted cautiously because outcomes are short term, and treatment was not randomly assigned. Only 5% of

nonsurgical patients received cognitive behavior therapy. Nonsurgical treatment that patients received was variable and mostly not compliant with major guidelines.

In addition, a large number of studies of adverse events, primarily large population-based observational studies, have been published in the past year. Examples of some of these studies include:

Spine (Phila Pa 1976). 2013 Dec 30. [Epub ahead of print]

Cerebral Vascular Accidents Following Lumbar Spine Fusion.

Marquez-Lara A, Nandyala SV, Fineberg SJ, Singh K.

Author information

Abstract

Study Design. Retrospective cohort.**Objective.** In order to determine the impact of a cerebral vascular accident (CVA) following lumbar spinal fusion, a population-based database was analyzed to identify the incidence, potential risk factors, hospital resource utilization, and the early postoperative outcomes.**Summary of Background.** A lumbar fusion (LF) is an effective surgical procedure to treat lumbar degenerative pathology. Although rare, a CVA can be a catastrophic event following a LF.**Methods.** The Nationwide Inpatient Sample (NIS) database was queried from 2002-2011. Patients undergoing an elective anterior lumbar fusion (ALF), a posterior lumbar fusion (PLF), or a combined anterior-posterior lumbar fusion (APLF) were separated into sub-cohorts. Patients with a documented postoperative CVA were identified. Patient demographics, comorbidities (CCI), length of stay (LOS), costs, early postoperative outcomes, and mortality were assessed. Statistical analysis involved T-tests, Chi-Square analysis, and binary logistic regression with $p < 0.001$ denoting significance.**Results.** A total of 264,891 LFs were identified between 2002-2011 of which 340 (1.3 per 1,000) developed a postoperative CVA. Patients with a CVA were significantly older and demonstrated a greater comorbidity burden (CCI). Patients with a CVA incurred a significantly greater LOS, total hospital costs (\$41,454 vs \$25,885), and a greater mortality rate (73.7 vs 0.8 per 1,000 patients). Regression analysis demonstrated that age > 65 years and a history of neurologic disorders, paralysis, congestive heart failure (CHF), or electrolyte imbalance were associated with an increased risk of a postoperative CVA.**Conclusions.** Patients who developed a postoperative CVA demonstrated a significantly greater incidence of postoperative complications, mortality, and total hospital costs. This study highlights important associated risk factors (e.g age > 65 , neurologic disorders, CHF) that may enable surgeons to identify high-risk patients prior to surgery. Further studies are warranted to characterize these risk

factors and to establish guidelines to mitigate the complications associated with a postoperative CVA.

Spine J. 2013 Oct 31. pii: S1529-9430(13)01610-0. doi: 10.1016/j.spinee.2013.10.015.
[Epub ahead of print]

Incidence and risk factors for postoperative ileus following anterior, posterior, and circumferential lumbar fusion.

Fineberg SJ1, Nandyala SV1, Kurd MF1, Marquez-Lara A1, Noureldin M1, Sankaranarayanan S1, Patel AA2, Oglesby M1, Singh K3.

Author information

Abstract

BACKGROUND CONTEXT:

Postoperative ileus is a known complication of surgery. The incidence and risk factors for ileus after lumbar fusion surgery is not well characterized.

PURPOSE:

To determine rates of postoperative ileus, a population-based database was analyzed to identify incidence, mortality, and risk factors associated with anterior (ALF), posterior (PLF), and combined anterior/posterior (APLF) lumbar fusions.

STUDY DESIGN:

This was a retrospective database analysis.

PATIENT SAMPLE:

The sample consisted of 220,522 patients from the Nationwide Inpatient Sample (NIS) database.

OUTCOME MEASURES:

Outcome measures were incidence of postoperative ileus, length of stay (LOS), in-hospital costs, and mortality.

METHODS:

Data from the NIS were obtained from 2002 to 2009. Patients undergoing ALF, PLF, and APLF for degenerative pathologies were identified and the incidence of postoperative ileus was assessed. Patient demographics, Charlson comorbidity index (CCI), LOS, costs, and mortality were assessed. SPSS v.20 was used to detect statistical differences between groups and perform logistic regression analyses to identify independent predictors of postoperative ileus. A p value less than .001 denoted significance.

RESULTS:

A total of 220,522 lumbar fusions were identified in the United States from 2002 to 2009. There were 19,762 ALFs, 182,801 PLFs, and 17,959 APLFs. The incidence of postoperative ileus was increased in ALFs over PLFs (74.9 vs. 26.0 per 1,000; $p < .001$). Within PLF and APLF groups, CCI scores were increased in the presence of postoperative ileus ($p < .001$). Across cohorts, patients with

postoperative ileus demonstrated greater LOS and costs ($p < .001$). PLF-treated patients with postoperative ileus demonstrated increased mortality ($p < .001$). Independent predictors of postoperative ileus included male gender, 3+ fusion levels, alcohol abuse, anemia, fluid/electrolyte disorders, and weight loss ($p < .001$).

CONCLUSIONS:

The results of our study demonstrate increased incidence of postoperative ileus associated with anterior approaches for lumbar fusion. Across cohorts, postoperative ileus was associated with increased LOS and costs. To determine the mortality and resource use associated with postoperative ileus, we recommend preoperatively identifying and treating modifiable risk factors, especially when an anterior approach is used.

Finally, neither more recent nor older studies on **specific outcomes of lumbar fusion in workers compensation** have been included in either the original HTA nor in the Hayes review as the AMDG initially requested. The Harris et al study in JAMA in 2005 provides compelling data on this issue across all procedures:

JAMA. 2005 Apr 6;293(13):1644-52.

Association between compensation status and outcome after surgery: a meta-analysis.

Harris J, Mulford J, Solomon M, van Gelder JM, Young J.

Author information

Abstract

CONTEXT:

Compensation, whether through workers' compensation or through litigation, has been associated with poor outcome after surgery; however, this association has not been examined by meta-analysis.

OBJECTIVE:

To investigate the association between compensation status and outcome after surgery.

DATA SOURCES:

We searched MEDLINE (1966-2003), EMBASE (1980-2003), CINAHL, the Cochrane Controlled Trials Register, and reference lists of retrieved articles and textbooks, and we contacted experts in the field.

STUDY SELECTION:

The review included any trial of surgical intervention in which compensation status was reported and results were compared according to that status. No restrictions were placed on study design, language, or publication date. Studies were selected by 2 unblinded independent reviewers.

DATA EXTRACTION:

Two reviewers independently extracted data on study type, study quality, surgical procedure, outcome, country of origin, length and completeness of follow-up, and compensation type.

DATA SYNTHESIS:

Two hundred eleven studies satisfied the inclusion criteria. Of these, 175 stated that the presence of compensation (workers' compensation with or without litigation) was associated with a worse outcome, 35 found no difference or did not describe a difference, and 1 described a benefit associated with compensation. A meta-analysis of 129 studies with available data (n = 20,498 patients) revealed the summary odds ratio for an unsatisfactory outcome in compensated patients to be 3.79 (95% confidence interval, 3.28-4.37 by random-effects model). Grouping studies by country, procedure, length of follow-up, completeness of follow-up, study type, and type of compensation showed the association to be consistent for all subgroups.

CONCLUSIONS:

Compensation status is associated with poor outcome after surgery. This effect is significant, clinically important, and consistent. Because data were obtained from observational studies and were not homogeneous, the summary effect should be interpreted with caution. Compensation status should be considered a potential confounder in all studies of surgical intervention. Determination of the mechanism for this association requires further study.

Finally, more recent data on **morbidity after lumbar fusion among WA state payers** has recently been published and should be included as a benchmark of morbidity rates:

Health Serv Res. 2013 Feb;48(1):1-25. doi: 10.1111/j.1475-6773.2012.01434.x. Epub 2012 Jun 20.

Hospital and surgeon variation in complications and repeat surgery following incident lumbar fusion for common degenerative diagnoses.

Martin BI, Mirza SK, Franklin GM, Lurie JD, MacKenzie TA, Deyo RA.

Author information**Abstract****OBJECTIVE:**

To identify factors that account for variation in complication rates across hospitals and surgeons performing lumbar spinal fusion surgery.

DATA SOURCES:

Discharge registry including all nonfederal hospitals in Washington State from 2004 to 2007.

STUDY DESIGN:

We identified adults (n = 6,091) undergoing an initial inpatient lumbar fusion for degenerative conditions. We identified whether each patient had a subsequent complication within 90 days. Logistic regression models with hospital and surgeon random effects were used to examine complications, controlling for patient characteristics and comorbidity.

PRINCIPAL FINDINGS:

Complications within 90 days of a fusion occurred in 4.8 percent of patients, and 2.2 percent had a reoperation. Hospital effects accounted for 8.8 percent of the total variability, and surgeon effects account for 14.4 percent. Surgeon factors account for 54.5 percent of the variation in hospital reoperation rates, and 47.2 percent of the variation in hospital complication rates. The discretionary use of operative features, such as the inclusion of bone morphogenetic proteins, accounted for 30 and 50 percent of the variation in surgeons' reoperation and complication rates, respectively.

CONCLUSIONS:

To improve the safety of lumbar spinal fusion surgery, quality improvement efforts that focus on surgeons' discretionary use of operative techniques may be more effective than those that target hospitals.

Conclusion: The AMDG strongly recommends re-review of lumbar fusion, and recommends re-specification of the original key questions to more accurately reflect the types of patients receiving lumbar fusion in the public insurance programs.

Spinal Cord Stimulation Assessing Signals for Update

Provided by:



Spectrum Research, Inc.

Prepared by:

Robin Hashimoto, PhD
Katie Moran, BS
Joseph R. Dettori, PhD, MPH
Jan 9, 2014

Previous Coverage Decision

A Comparative Effectiveness Review (CER) titled: SPINAL CORD STIMULATION, was originally released in July 2010 by the Health Technology Clinical Committee and summarized below.

Health Technology Clinical Committee

Findings and Coverage Decision
 Topic: SPINAL CORD STIMULATION
 Meeting Date: August 20, 2010
 Final Adoption: October 22, 2010

HTCC Coverage Determination

Spinal Cord Stimulation for chronic neuropathic pain is **not a covered benefit**.

HTCC Reimbursement Determination

- **Limitations of Coverage**
Not Applicable
- **Non-Covered Indicators**
Not applicable

Health Technology Background

The Spinal Cord Stimulation topic was selected and published in December 2009 to undergo an evidence review process. Spinal Cord Stimulation (SCS) is an alternative treatment proposed for patients with chronic neuropathic pain who have not responded to conventional therapies such as medication, physical and/or psychological therapy, and in some case, re-operation. Current best evidence is available primarily from four trials on 375 patients; which are rated at a Level 1 or 2 (good quality), which is a better level of evidence than some interventions. However, total patient sample size is small, comparators were weak or inappropriate, reported outcomes are mostly subjective and not consistently reported, industry funding and management may have an impact, and no trial included a sham stimulation/procedure arm. The overall body of evidence was inconsistent, with several trials showing benefits on some outcomes at generally shorter follow up periods and others showing no difference. SCS is an implanted, long term treatment, but no evidence exists on either long term efficacy or safety.

The committee agreed that SCS is less safe than alternatives, is an invasive procedure, and has many adverse events. While conventional medical management is not invasive, so would generally have a lower risk profile, re-operation is also a comparator and had less complications. SCS device related complications can be serious and include dural punctures, amplitude by bodily movements; paresthesia in other body parts, pain, disturbed urination, lead fracture, loss of effect, infection. Indications for SCS (FDA): Chronic intractable pain in the trunk and/or limbs including unilateral or bilateral pain associated with FBSS and intractable low back and leg pain, and for some devices: CRPS, radicular pain syndrome or radiculopathies resulting in pain, post-laminectomy pain, unsuccessful disc surgery, degenerative disc disease or herniated disc pain refractory to conservative or surgical interventions, peripheral causalgia, epidural fibrosis, arachnoiditis or lumbar adhesive arachnoiditis, and multiple back surgeries. Potential patients should undergo a period of trial stimulation prior to permanent SCS implantation. Contraindications for SCS (FDA): Failed trial stimulation due to ineffective pain relief; poor surgical risks; pregnancy; active general infections or multiple illnesses; inability to operate the SCS system; and cardiac pacemakers (with specific exceptions and precautions) or cardioverter defibrillators.

In June 2010, the HTA posted a draft and then followed with a final report from a contracted research organization that reviewed publicly submitted information; searched, summarized, and evaluated trials, articles, and other evidence about the topic. The comprehensive, public and peer reviewed Spinal Cord Stimulation report is 164 pages, and identified a relatively large amount of literature.

An independent group of eleven clinicians who practice medicine locally meet in public to decide whether state agencies should pay for the health technology based on whether the evidence

report and other presented information shows it is safe, effective and has value. The committee met on August 20, reviewed the report, including peer and public feedback, and heard public and agency comments. Meeting minutes detailing the discussion are available through the HTA program or online at <http://www.hta.hca.wa.gov> under the committee section.

Committee Conclusions

Having made findings as to the most significant and relevant evidence regarding health outcomes, key factors and identified evidence related to those factors, primarily based on the evidence based technology assessment report, the committee concludes:

(1) Evidence availability and technology features

The committee concludes that the best available evidence on Spinal Cord Stimulation has been collected and summarized.

- Spinal Cord Stimulation (SCS) is an alternative treatment proposed for patients with chronic neuropathic pain who have not responded to conventional therapies such as medication, physical and/or psychological therapy, and in some case, re-operation.
- Current best evidence is available primarily from four trials on 375 patients; which are rated at a Level 1 or 2 (good quality), which is a better level of evidence than some interventions. However, total patient sample size is small, comparators were weak or inappropriate, reported outcomes are mostly subjective and not consistently reported, industry funding and management may have an impact, and no trial included a sham stimulation/procedure arm. The overall body of evidence was inconsistent, with several trials showing benefits on some outcomes at generally shorter follow up periods and others showing no difference.
- SCS is an implanted, long term treatment, but no evidence exists on either long term efficacy or safety.

(2) Is it safe?

The committee concludes that the comprehensive evidence indicates that Spinal Cord Stimulation is less safe than alternative treatments. Key factors to the committee's conclusion included:

- The committee agreed that SCS is less safe than alternatives, is an invasive procedure, and has many adverse events. While conventional medical management is not invasive, so would generally have a lower risk profile, re-operation is also a comparator and had less complications. SCS device related complications can be serious and include dural punctures, amplitude by bodily movements; paresthesia in other body parts, pain, disturbed urination, lead fracture, loss of effect, infection.
- The committee agreed that safety was a significant factor: the number of trial reported complications ranged from 8 to 100%. Device related complication requiring revision ranged from 25% to 38% of patients in short term and 42% to 60% in up to 5 years (not including 54% of patients undergoing pulse generator replacements due to battery life).
- The committee agreed that there were currently no reported mortality rates, but that the FDA data was not available and the small sample size is likely underpowered to detect.
- The committee agreed that the removal rate could be considered an efficacy or safety issue, but the rates ranging from 4% to 17% were concerning, especially considering that trial stimulation is done first on all patients.

(3) Is it effective?

The majority of the committee concludes that the comprehensive evidence about Spinal Cord Stimulation effectiveness is unproven.

- The committee agreed that the studies had serious limitations in design, low patient sample sizes, and weak or inadequate comparators. Additionally, placebo effects of a new intervention for patients with chronic pain who have already failed multiple therapies is a serious concern and no study involved sham stimulation or procedures and outcome measures were generally subjective.
- The committee found that evidence overall on important patient outcomes was limited. For all

outcomes, there is no evidence of longer term improvement, particularly important when there are significant risks (including 1/3 revision and high removal rate) and the device is intended for permanent implant.

- Given the serious limitations of the studies, the committee agreed that, at best, weak evidence exists that SCS may provide temporary improvement of pain in some patients, but there is no evidence of mid or long term pain improvement.
- While pain is a critical patient outcome, evidence about other important patient outcomes was either not available or not consistent with the pain findings.
 - For instance, for reduction in pain medication in short term: Kumar and Turner found no difference, while North found SCS patients did have reduction.
 - For functional improvements, 1 trial found short term functional improvement, but 2 others did not; and there was no reliable evidence of functional improvement at mid (or long) term.
- For all other outcomes, including improvement in quality of life, there is no reliable evidence of effect.

(4) Evidence about the technology's special populations, patient characteristics and adjunct treatment

The committee agreed that no compelling evidence exists to differentiate sub groups or special populations.

- The committee agreed with the evidence based report that there is inadequate evidence to identify characteristics that either enhance or reduce the efficacy of SCS such as age, sex, workers' compensation or other disability payments, duration of pain, pain intensity, time since first lumbar surgery, number of prior operations for pain, pain location, laterality of pain, allodynia or hypoesthesia at baseline, McGill Pain Questionnaire or the Minnesota Multiphasic Personality Inventory (MMPI)

(5) Is the technology cost-effective?

- The committee concludes that SCS is unproven to be cost effective.
- The committee agreed that the cost of SCS is substantial, averaging \$27,000 per patient.
- The committee agreed that overall value cannot be ascertained without evidence of net benefit of effectiveness and reduced harm. Reliable cost-effectiveness analysis cannot be performed.

Committee Decision

Based on the deliberations of key health outcomes, the committee decided that it had the most complete information: a comprehensive and current evidence report, public comments, and agency and state utilization information. The committee concluded that the current evidence on Spinal Cord Stimulation demonstrates that there isn't sufficient evidence to cover the use of Spinal Cord Stimulation for chronic neuropathic pain. The committee considered all the evidence and gave greatest weight to the evidence it determined, based on objective factors, to be the most valid and reliable. Based on these findings, the committee voted 8 to 1 to not cover Spinal Cord Stimulation.

The committee reviewed the Clinical guidelines and Medicare decision. The Medicare decision was did not cite evidence and was decided prior to any of the studies reviewed by the committee. The guidelines recommendations conflict and not all have reviewed the latest trials included in this report.

1. Purpose of Report

The purpose of this literature update is to determine whether or not there is sufficient evidence published after the original report to conduct a re-review of this technology.

2. Methods

2.1 Literature Searches

We conducted a limited literature search for articles published between May 1, 2010 and December 6, 2013 using the identical search strategy used for the original report. This search included four main databases: PubMed, Medline, Cochrane Library, and EMBASE. Appendix A includes the search methodology for this topic.

2.2 Study selection

In general, we used the same inclusion and exclusion criteria as the original CER.

2.4 Compilation of Findings and Conclusions

For this assessment we abstracted the data from the included studies and constructed a demographics table, Table 3 (Appendix C). We also constructed a summary table that included the key questions, the original conclusions, new sources of evidence, new findings, and conclusions based on available signals, Table 2. To assess whether the conclusions might need updating, we used an algorithm based on a modification of the Ottawa method, Figure 2.

3. Results

3.1 Search

A systematic review was undertaken for articles published between May 1, 2010 and December 6, 2013. We used two search strategies to identify articles from MEDLINE, EMBASE and the Cochrane Library. We used key words to detect articles that used the terms “spinal cord stimulation”, “spinal cord stimulator”, or “spinal cord stimulation”. Among the articles describing the efficacy and/or safety of spinal cord stimulation, we evaluated the full text to determine if the studies met our inclusion criteria. Full text of potential articles meeting the inclusion criteria by both methods were reviewed by two independent investigators to obtain the final collection of included studies, Figure 1.

The literature search identified 213 titles. After title and abstract review, we further reviewed the full text of 22 journal articles. The remaining 191 titles were rejected because they were case reports, commentary, or did not include topics of interest. Among the 22 articles that went on to full text review, 17 were rejected because subjects did not meet the inclusion criteria and/or did not include a comparison of interest, Table 1. No new systematic reviews of relevant literature were identified. Of the five articles that were further reviewed, all five¹⁻⁵ were abstracted into an evidence table (Appendix C).

3.2 Identifying signals for re-review

Table 2 shows the original key questions, the conclusions of the original report, the new sources of evidence, the new findings, and the recommendations of Spectrum Research, Inc. (SRI) regarding the need for update.

4. Conclusions (Appendix B, Table 2)

4.1 Key Question 1: What is the evidence of efficacy and effectiveness of spinal cord stimulation?

Efficacy: All conclusions are still valid and this portion of the CER does not need updating.

Effectiveness: All conclusions are still valid and this portion of the CER does not need updating.

4.2 Key Question 2: What is the evidence of the safety of spinal cord stimulation?

All conclusions are still valid and this portion of the CER does not need updating.

4.3 Key Question 3: What is the evidence that spinal cord stimulation has differential efficacy or safety issues in sub populations?

All conclusions are still valid and this portion of the CER does not need updating.

4.4 Key Question 4: What is the evidence of cost implications and cost-effectiveness of spinal cord stimulators?

- This section of the report could be updated with the results of the cost-effectiveness analysis of the cohort of Washington State workers' compensation patients with FBSS (Hollingworth (2011)⁴.
- However, the addition of this analysis (which suggests that SCS is not cost-effective in this patient population compared with pain clinic or usual care) would not affect the coverage decision (SCS is not covered).

References:

1. Falowski SM, Celii A, Sestokas AK, Schwartz DM, Matsumoto C, Sharan A. Awake vs. asleep placement of spinal cord stimulators: a cohort analysis of complications associated with placement. *Neuromodulation* 2011;14:130-4; discussion 4-5.
2. Kumar K, Rizvi S, Bnurs SB. Spinal cord stimulation is effective in management of complex regional pain syndrome I: fact or fiction. *Neurosurgery* 2011;69:566-78; discussion 5578-80.
3. Wolter T, Kieselbach K. Cervical spinal cord stimulation: an analysis of 23 patients with long-term follow-up. *Pain Physician* 2012;15:203-12.
4. Hollingworth W, Turner JA, Welton NJ, Comstock BA, Deyo RA. Costs and cost-effectiveness of spinal cord stimulation (SCS) for failed back surgery syndrome: an observational study in a workers' compensation population. *Spine (Phila Pa 1976)* 2011;36:2076-83.
5. Kemler MA, Raphael JH, Bentley A, Taylor RS. The cost-effectiveness of spinal cord stimulation for complex regional pain syndrome. *Value Health* 2010;13:735-42.

Figure 1. Flow chart showing results of literature search

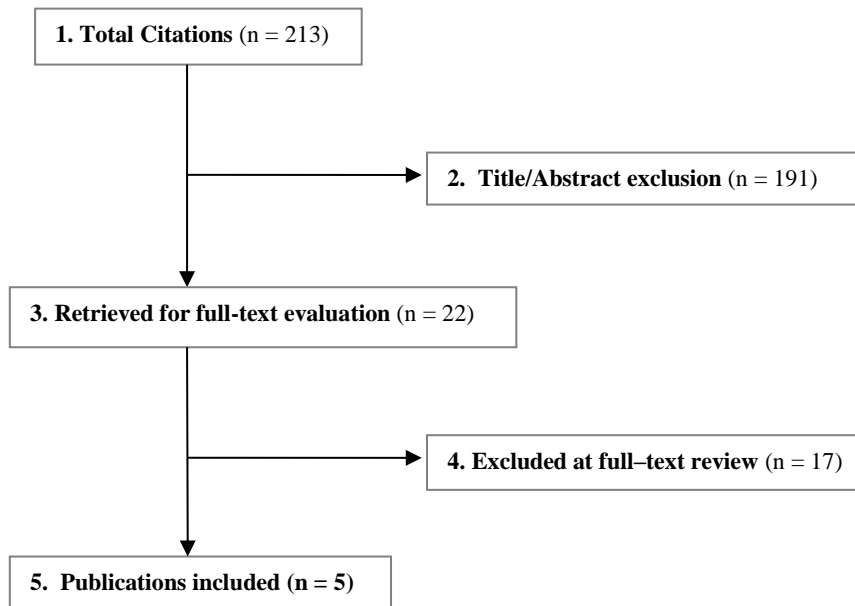
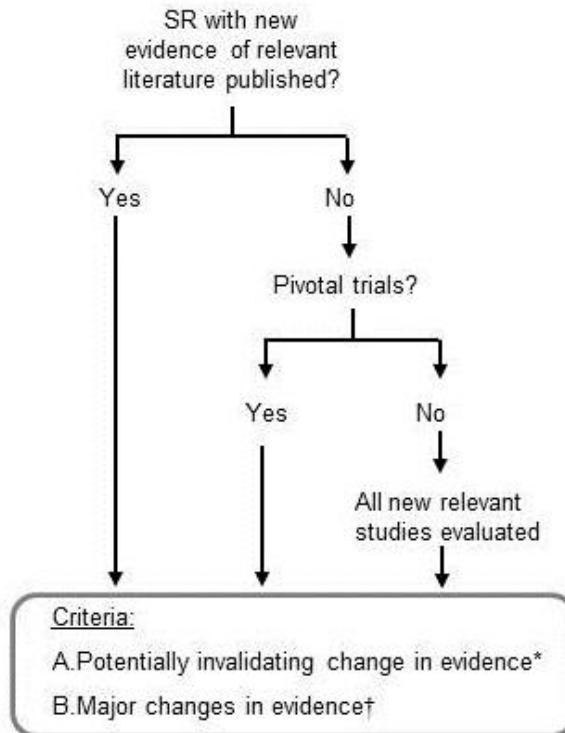


Figure 2. Algorithm of the Ottawa Method of Identifying Signals for SR Updates

Algorithm using a modified version of the Ottawa Method of identifying signals for SR updates



*A-1. Opposing findings: Pivotal trial or SR including at least one new trial that characterized the treatment in terms opposite to those used earlier

A-2. Substantial harm: Pivotal trial or SR whose results called into question the use of the treatment based on evidence of harm or that did not proscribe use entirely but did potentially affect clinical decision making

A-3. Superior new treatment: Pivotal trial or SR whose results identified another treatment as significantly superior to the one evaluated in the original review, based on efficacy or harm.

†B-1. Important changes in effectiveness short of "opposing findings"

B-2. Clinically important expansion of treatment

B-3. Clinically important caveat

B-4. Opposing findings from discordant meta-analysis or nonpivotal trial

Table 1. List of excluded articles after full-text review

| Study | Reason for Exclusion: |
|---|---|
| Systematic reviews | |
| Kelly GA, Blake C, Power CK, O'Keeffe D, Fullen BM. The impact of spinal cord stimulation on physical function and sleep quality in individuals with failed back surgery syndrome: a systematic review. <i>Eur J Pain</i> 2012;16:793-802. | Systematic review does not contain any studies published after the search period of the original HTA. |
| Krames ES, Monis S, Poree L, Deer T, Levy R. Using the SAFE principles when evaluating electrical stimulation therapies for the pain of failed back surgery syndrome. <i>Neuromodulation</i> 2011;14:299-311; discussion | Not a systematic review. |
| Levy R, Henderson J, Slavin K, et al. Incidence and avoidance of neurologic complications with paddle type spinal cord stimulation leads. <i>Neuromodulation</i> 2011;14:412-22; discussion 22. | Systematic review does not contain any relevant studies published after the search period of the original Washington State HTA. |
| Lihua P, Su M, Zejun Z, Ke W, Bennett MI. Spinal cord stimulation for cancer-related pain in adults. <i>Cochrane Database Syst Rev</i> 2013;2:CD009389 | Systematic review does not contain any relevant studies published after the search period of the original HTA (SR included 4 case series, two of which were published after the search of the original HTA, and neither of these reported on adverse events following SCS). |
| Mailis A, Taenzer P. Evidence-based guideline for neuropathic pain interventional treatments: spinal cord stimulation, intravenous infusions, epidural injections and nerve blocks. <i>Pain Res Manag</i> 2012;17:150-8. | Systematic review does not contain any relevant studies published after the search period of the original HTA. |
| Manchikanti L, Abdi S, Atluri S, et al. An update of comprehensive evidence-based guidelines for interventional techniques in chronic spinal pain. Part II: guidance and recommendations. <i>Pain Physician</i> 2013;16:S49-283. | Systematic review does not contain any relevant studies published after the search period of the original HTA. |
| Pluijms WA, Slangen R, Joosten EA, et al. Electrical spinal cord stimulation in painful diabetic polyneuropathy, a systematic review on treatment efficacy and safety. <i>Eur J Pain</i> 2011;15:783-8. | Systematic review does not contain any relevant studies published after the search period of the original Washington State HTA. |
| Poree L, Krames E, Pope J, Deer TR, Levy R, Schultz L. Spinal cord stimulation as treatment for complex regional pain syndrome should be considered earlier than last resort therapy. <i>Neuromodulation</i> 2013;16:12 | Not a systematic review. |
| Raff M, Melvill R, Coetsee G, Smuts J. Spinal cord stimulation for the management of pain: Recommendations for best clinical practice. <i>S Afr Med J</i> 2013;103:423-30. | Search period of systematic review portion does not extend beyond the search period of the original Washington State HTA. |
| Sparkes E, Raphael JH, Duarte RV, LeMarchand K, Jackson C, Ashford RL. A systematic literature review of psychological characteristics as determinants of outcome for spinal cord stimulation therapy. <i>Pain</i> 2010;150:284-9. | Search period does not extend beyond the search period of the original Washington State HTA. |
| van Wijck AJ, Wallace M, Mekhail N, van Kleef M. Evidence-based interventional pain medicine according to clinical diagnoses. 17. Herpes zoster and post-herpetic neuralgia. <i>Pain Pract</i> 2011;11:88-97. | Not a systematic review. |
| KQ1 | |
| North RB, Kumar K, Wallace MS, et al. Spinal cord stimulation versus re-operation in patients with failed back surgery syndrome: an international multicenter randomized controlled trial (EVIDENCE study). <i>Neuromodulation</i> 2011;14:330-5; discussion 5-6. | Study protocol only; excluded as the trial has been terminated due to slow enrollment: http://clinicaltrials.gov/show/NCT01036529 |

| | |
|--|--|
| KQ2 | |
| Kim DD, Vakharya R, Kroll HR, Shuster A. Rates of lead migration and stimulation loss in spinal cord stimulation: a retrospective comparison of laminotomy versus percutaneous implantation. <i>Pain Physician</i> 2011;14:513-24. | Case series with less than five years' follow-up. |
| Mekhail NA, Mathews M, Nageeb F, Guirguis M, Mekhail MN, Cheng J. Retrospective review of 707 cases of spinal cord stimulation: indications and complications. <i>Pain Pract</i> 2011;11:148-53. | Case series with less than five years' follow-up. |
| Zan E, Kurt KN, Yousem DM, Christo PJ. Spinal cord stimulators: typical positioning and postsurgical complications. <i>AJR Am J Roentgenol</i> 2011;196:437-45. | Case series with length of follow-up not reported. |
| KQ3 | |
| Williams KA, Gonzalez-Fernandez M, Hamzehzadeh S, et al. A multi-center analysis evaluating factors associated with spinal cord stimulation outcome in chronic pain patients. <i>Pain Med</i> 2011;12:1142-53. | Case series, so LoE of IV (not I or II as specified in the inclusion criteria of the original HTA). |
| KQ4 | |
| Taylor RS, Ryan J, O'Donnell R, Eldabe S, Kumar K, North RB. The cost-effectiveness of spinal cord stimulation in the treatment of failed back surgery syndrome. <i>Clin J Pain</i> 2010;26:463-9. | Article provides "previously unavailable details of the NICE cost-effectiveness analysis", which was included in the original HTA. This article reaches similar conclusions as those of the NICE analysis, that is, that SCS is cost-effective compared to CMM or re-operation in FBSS patients. |

Appendix A.

The detailed strategy below is presented in Medline and EMBASE syntax.

Search Strategy

(May 2010 – December 6, 2013)

Limited to English language, human population

Database: MEDLINE

| | |
|----|---|
| 1. | "Spinal cord stimulation" OR "Spinal cord stimulation"[MeSH] OR "spinal cord stimulator" OR "spinal cord stimulators" |
| 2. | #1 NOT "Case Reports"[Publication Type] |

Database: EMBASE

| |
|--|
| 'spinal cord stimulation'/exp OR 'spinal cord stimulator'/exp AND [humans]/lim AND [English]/lim AND [abstracts]/lim AND [5-1-2013]/sd NOT [12-1-2013]/sd AND [2010-2014]/py |
|--|

Parallel strategies were used to search the Cochrane Library and others listed below. Keyword searches were conducted in the other listed resources.

Electronic Database Searches

The following databases have been searched for relevant information:

Agency for Healthcare Research and Quality (AHRQ)
 Cumulative Index to Nursing and Allied Health (CINAHL)
 Cochrane Database of Systematic Reviews (through 2009, Issue 2)
 Cochrane Registry of Clinical Trials (CENTRAL) (through 2009, Issue 2)
 Cochrane Review Methodology Database (through 2009, Issue 2)
 Computer Retrieval of Information on Scientific Projects (CRISP)
 Database of Reviews of Effectiveness (Cochrane Library) (through 2009, Issue 2)
 EMBASE (1985 through July 23, 2009)
 PubMed (1975 through July 23, 2009)
 Informational Network of Agencies for Health Technology Assessment (INAHTA)
 NHS Economic Evaluation Database (Cochrane Library through 2009, Issue 2)
 HSTAT (Health Services/Technology Assessment Text)
 EconLIT

Additional Economics, Clinical Guideline and Gray Literature Databases

AHRQ- Healthcare Cost and Utilization Project
 Canadian Agency for Drugs and Technologies in Health
 Centers for Medicare and Medicaid Services (CMS)
 Food and Drug Administration (FDA)
 Google
 Institute for Clinical Systems Improvement (ICSI)
 National Guideline Clearinghouse

Appendix B

Table 2. Spinal Cord Stimulation Summary Table

| Conclusions from CER Executive Summary | New Sources of Evidence | New Findings | Conclusion from SRI |
|---|-------------------------|--------------|--|
| Key Question 1: What is the evidence of efficacy and effectiveness of spinal cord stimulation? | | | |
| <p>1. a) Efficacy (Short-term, <5 years):</p> <ul style="list-style-type: none"> • Pain, perceived effect of treatment/patient satisfaction: There is moderate evidence from three small randomized controlled trials that SCS is superior to conventional therapies (CMM, physical therapy or re-operation) in patients with chronic neuropathic pain during the first 2–3 years with respect to patient reported outcomes of pain, and perceived effect of treatment/patient satisfaction. In the only RCT that measured outcomes for a longer period of time, the benefit of SCS decreased over time and was not significantly different than controls for leg pain after 3 years of treatment (see mid-term below). • Function, quality of life: The effect on quality of life outcomes is less clear with one RCT reporting substantial benefit of SCS compared with CMM at 6 months follow-up, while another study found quality of life outcomes to be similar between SCS + physical therapy and physical therapy alone at 2 years follow-up. Similarly, function as measured by the Oswestry Disability Index score was better in the SCS group at 6 months versus CMM in one study but the ability to perform daily activities after 3 years was not different in a second study. The strength of this evidence is low. <p>b) Efficacy (Mid-term, 5-10 years):</p> <ul style="list-style-type: none"> • Pain, quality of life, perceived effect of treatment: There is low evidence from one small randomized controlled trial that SCS is no different from conventional therapy (physical therapy) in patients with chronic neuropathic pain 5-10 years following implant with respect to pain, quality of life, and patient-reported global perceived effect. <p>c) Efficacy (Long-term, ≥10 years):</p> <ul style="list-style-type: none"> • There are no data available to assess long-term efficacy. <p>2. a) Effectiveness (Short-term, <5 years):</p> <ul style="list-style-type: none"> • Composite of pain, function, and opioid use: One prospective cohort study on workers’ compensation patients reported similar success on a composite score that includes pain, function and opioid use between SCS | None | None | <ul style="list-style-type: none"> • This section of the report is still valid and does not need updating |

| Conclusions from CER Executive Summary | New Sources of Evidence | New Findings | Conclusion from SRI |
|---|--|--|--|
| <p>and either Pain Clinic or Usual Care treatment groups. There was a modest improvement in leg pain in the SCS group compared with the control groups at 6 months follow-up but this did not persist at the 12 month or 24 month evaluation.</p> <p>b) Effectiveness (Mid- and long-term, ≥5 years):</p> <ul style="list-style-type: none"> There are no data available to assess mid- or long-term effectiveness. | | | |
| <p>Key Question 2: What is the evidence of the safety of spinal cord stimulation?</p> | | | |
| <p>1. Revision</p> <ul style="list-style-type: none"> There is high evidence from three randomized controlled trials, one prospective comparative cohort study and six case series that revision of SCS components is not uncommon. Overall short-term revision rates ranged from 12–38% of patients. Mid-term revision rates were 42% in one RCT and 60% in one case series. Reasons for revision include electrode repositioning or replacement, generator revision or replacement, revision of the connecting cable, and total removal and replacement of the system due to infection. There are no long-term data available. <p>2. Other SCS-related side effects</p> <ul style="list-style-type: none"> Side effects reported varied widely among studies and included infection, change in amplitude by bodily movements, paresthesia in other body parts, pain/irritation from the pulse generator, transient neurological defects, severe wound-related pain at the stimulator implantation site, cerebrospinal fluid leak, and subcutaneous hematoma. The rate of side effects could not be determined from the papers reviewed; however, one RCT reported that all patients experienced at least one side effect. <p>3. Mortality</p> <ul style="list-style-type: none"> There is high evidence that the rate of mortality due to SCS is low. Among the four comparative studies, 2 deaths were reported in patients receiving SCS (2/139); one as a result of a cardiac event six months following SCS implantation, and the cause of one was not reported. No deaths were recorded in the control groups during the same time period (0/179). Two additional deaths were identified in three case series with five year follow-up; one from a cerebrovascular accident in a patient implanted for cardiac ischemic pain, one as a result of suicide. No death was attributed to SCS; however one patient nearly died as a result of complications that arose following trial stimulation. | <p><u>3 case series:</u> Falowski (2011)¹ Kumar (2011)² Wolter (2012)³</p> | <ul style="list-style-type: none"> There is very low evidence from three case series¹⁻³ of a total of 305 patients that revision rates from device failure, injection, device/electrode repositioning, electrode fracture, electrode replacement, battery end of life, or pain at the implantation site) range from 14% to 50% of patients. The mean length follow-up was >5 to 7.3 years. There is very low evidence from one small case series of 25 patients that there were no bleeding complications; this series had a mean length follow-up of 7.3 years.² There is very low evidence from one small case series of 18 patients that there were no severe neurological deficits; this series had a mean length follow-up of 5.8 years.³ There is very low evidence from one small case series of 18 patients that 22% of patients had pain at the implant site; this series had a mean length follow-up of 5.8 years.³ | <ul style="list-style-type: none"> This section of the report is still valid and does not need updating |

| Conclusions from CER Executive Summary | New Sources of Evidence | New Findings | Conclusion from SRI |
|---|-------------------------|--------------|--|
| Key Question 3: What is the evidence that spinal cord stimulation has differential efficacy or safety issues in sub populations? | | | |
| <p>1. Age</p> <ul style="list-style-type: none"> There is conflicting evidence whether patient age at baseline is associated with outcome. Two studies found that age did not correlate with either pain relief or success (combination of pain relief and patient satisfaction), while one study found that younger age was correlated with pain relief of at least 50%. One of these studies also reported no correlation between age and SF-36 or GPE scores. <p>2. Sex</p> <ul style="list-style-type: none"> There are mixed results regarding whether patient sex is associated with outcome following SCS. Three studies found that sex was not associated with pain relief, one showed no correlation between sex and SF-36 or GPE scores. In contrast, one study found that females had a significantly higher rate of success (pain relief and patient satisfaction), improved function and activity, and decreased medication usage at five years compared with males. <p>3. Workers' compensation or other disability payments</p> <ul style="list-style-type: none"> One prospective study suggests that whether patients receive workers' compensation/other disability payments or no compensation has no effect on pain relief among patients receiving SCS. Another prospective study found that among patients on workers' compensation, successful outcomes of pain relief, improved function and reduced opioid use was similar between SCS and two control treatment groups. The percentages of success were low in all groups. <p>4. Duration of pain</p> <ul style="list-style-type: none"> There is moderate evidence from three cohort studies that duration of pain prior to SCS implantation is not associated with pain relief or success within the first year after implantation. <p>5. Pain intensity</p> <ul style="list-style-type: none"> There is low evidence from one cohort study to suggest that pain intensity at baseline is not associated with success. <p>6. Time since first lumbar surgery</p> <ul style="list-style-type: none"> There is low evidence from one cohort study to suggest that time since first lumbar surgery is not predictive of success. | None | None | <ul style="list-style-type: none"> This section of the report is still valid and does not need updating |

| Conclusions from CER Executive Summary | New Sources of Evidence | New Findings | Conclusion from SRI |
|---|-------------------------|--------------|---------------------|
| <p>7. Number of prior surgeries for pain</p> <ul style="list-style-type: none"> • There is moderate evidence from two cohort studies to suggest that the number of prior of operations for pain is not associated with pain relief (or success). One study additionally found no correlation between prior operations for pain and function/activity/medication usage at five years. <p>8. Pain location</p> <ul style="list-style-type: none"> • There is low evidence from four cohort studies that pain location does not affect outcomes. <p>9. Laterality of pain</p> <ul style="list-style-type: none"> • There is low evidence from one cohort study on FBSS patients with open workers’ compensation claims that patients with unilateral pain have better pain relief and functional outcomes (as measured by the RDQ) at 12 months compared with patients with bilateral pain. <p>10. Allodynia or hypoesthesia at baseline</p> <ul style="list-style-type: none"> • There is low evidence from one cohort study that the presence of allodynia at baseline negatively correlates with success at one year, while the presence of hypoesthesia at baseline was not predictive of success. <p>11. McGill Pain Questionnaire</p> <ul style="list-style-type: none"> • There is conflicting evidence from two studies that the McGill Pain Questionnaire is associated with pain relief or success at follow-up with conflicting results. One study found an association between the evaluative subscale while the other study found no association with any subscale and outcome. <p>12. Minnesota Multiphasic Personality Inventory (MMPI)</p> <ul style="list-style-type: none"> • There is conflicting evidence from two studies that the MMPI is associated with pain relief or success at follow-up with conflicting results. One study found an association between the depression subscale while the other study found no association with any subscale and outcome. <p>13. SF-36 Mental Health scores</p> <ul style="list-style-type: none"> • There is low evidence from one cohort study on FBSS patients with open workers’ compensation claims that patients with baseline SF-36 Mental Health scores in the top third have better pain relief and functional outcomes (as measured by the RDQ) at 12 months than do those patients who scored in the bottom third at baseline. | | | |

| Conclusions from CER Executive Summary | New Sources of Evidence | New Findings | Conclusion from SRI |
|---|--|---|---|
| Key Question 4: What is the evidence of cost implications and cost-effectiveness of spinal cord stimulation? | | | |
| <p>Cost Effectiveness</p> <ul style="list-style-type: none"> There is moderate evidence from three complete economic evaluations that in the short-term, SCS is associated with improved outcomes and increased costs compared with CMM and/or re-operation for the treatment of neuropathic pain. In the long-term, SCS appears to be dominant over the control treatments; however, only one study included in this assessment was conducted in a U.S. setting. More specifically, we found that there is some evidence that SCS is cost-effective at moderate (<\$20,000) incremental cost effectiveness ratio (ICER) levels compared with CMM or re-operation, and that SCS cost-effectiveness increases and may be dominant over time compared with control treatments (i.e., CMM or re-operation) assuming device longevity of 4 years and at least a 30% pain threshold criteria. However, the assumption of continued efficacy past 3 years is questionable from the only RCT reporting pain 5-10 years after implantation. Furthermore, only one study was conducted in a US setting. | <p><u>2 cost-effectiveness analyses:</u> Hollingworth (2011)⁴ Kemler (2010)⁵</p> | <ul style="list-style-type: none"> Hollingworth (2011)⁴ evaluated the cost-effectiveness of SCS compared to pain clinic or usual care in a cohort of Washington State workers' compensation patients with FBSS. SCS was not cost-effective compared with usual care or pain clinic treatment. Kemler (2010) conducted a re-analysis of the data used in the NHS/NICE cost-effectiveness analysis of CRPS I patients, though the update was not published by NHS/NICE. The NHS/NICE analysis was included in the original HTA. This analysis arrived at similar conclusions as the original NHS/NICE evaluation, where SCS plus CMM is cost-effective compared to CMM alone. <ul style="list-style-type: none"> The ICER of £3562 per QALY for SCS compared with CMM was lower than that reported in the NHS/NICE report (£25,095 per QALY) (and included in the original HTA). | <ul style="list-style-type: none"> This section of the report could be updated with the results of the cost-effectiveness analysis of the cohort of Washington State workers' compensation patients with FBSS. However, the addition of this analysis (which suggests that SCS is not cost-effective in this patient population compared with pain clinic or usual care) would not affect the coverage decision (SCS is not covered). |

Appendix C. Demographic table

Table 3. Spinal cord stimulation studies demographic table

| Author (Year) Study type | Key Question | Demographics | Results | Conclusion |
|----------------------------------|--------------|--|---|--|
| Falowski 2011 Case series | KQ2 | <p>N = 259 (167 new device implantations, 220 re-operations for device failure, device repositioning, battery end of life, infection) Male: NR Age: NR F/U: >5 years (details NR)</p> <p><u>Diagnosis:</u> Neuropathic pain</p> <p><u>Intervention:</u> SCS implanted while patient awake (76 first-time procedures) or under general anesthesia (91 first-time procedures)</p> | <p>Safety:</p> <ul style="list-style-type: none"> • <u>Device failure</u> (including multiple surgeries); f/u: > 5 years (<i>Failure</i>: any re-operation secondary to a traumatic break in the SCS system, a device malfunction requiring re-exploration, or a device removal secondary to lack of efficacy) <ul style="list-style-type: none"> • Range: 14.9% - 29.7% of procedures <ul style="list-style-type: none"> • SCS implanted while patient awake: 29.7% • SCS implanted while patient asleep: 14.9% • <u>Infection requiring device explantation</u>; f/u: NR (details NR) <ul style="list-style-type: none"> • Range: 4.5% - 5.7% of procedures <ul style="list-style-type: none"> • SCS implanted while patient awake: 4.5% • SCS implanted while patient asleep: 5.7% • <u>Electrode repositioning</u>; f/u: NR (details NR) <ul style="list-style-type: none"> • Range: 14.9 – 17.9% of procedures <ul style="list-style-type: none"> • SCS implanted while patient awake: 17.9% • SCS implanted while patient asleep: 14.9% | <p>Safety:</p> <ul style="list-style-type: none"> • <u>Device failure</u> occurred in 14.9% to 29.7% of procedures (exact number not calculable) • <u>Infection requiring device explantation</u> occurred in 4.5% to 5.7% of procedures (exact number not calculable) • <u>Electrode repositioning</u> occurred in 4.5% to 5.7% of procedures (exact number not calculable) |
| Kumar (2011) Case series | KQ2 | <p>N = 28 Male: 43% Age: 51 (32-82) years F/U: 7.3 (1.5-19.6) years</p> <p><u>Diagnosis:</u> CRPS I</p> <p><u>Intervention:</u> SCS</p> | <p>Safety:</p> <ul style="list-style-type: none"> • <u>Device repositioning</u>: 20% (5/25) • <u>Electrode fracture</u>: 5% (1/25) • <u>Electrode repositioning</u>: 20% (5/25) • <u>Electrode replacement</u>: 44% (11/25) • <u>Battery end of life</u>: 40% (10/25) • <u>Hardware malfunction</u>: 0% (0/25) • <u>Infection requiring explantation and re-implantation</u>: 5% (1/25) (occurred 3 times) • <u>Bleeding</u>: 0% (0/25) | <ul style="list-style-type: none"> • This small case series suggested that when followed in the long-term (mean follow-up: 7.3 years), patients have a relatively high rate of hardware complications requiring re-operation (ranging from 0% to 44% of patients). The incidence of infection and bleeding were low (5% and 0%, respectively). |
| Wolter (2012) | KQ2 | N = 18 | Safety: | <ul style="list-style-type: none"> • This small case series reported that |

| Author (Year) Study type | Key Question | Demographics | Results | Conclusion |
|---|--------------|--|---|--|
| Case series | | <p>Male: 3% Age: 54 (34-78) years F/U: 5.8 (0.4-21) years</p> <p><u>Diagnosis:</u> Various types of cervical neuropathic pain</p> <p><u>Intervention:</u> SCS</p> | <ul style="list-style-type: none"> • Total “unscheduled” re-operations: 50% (9/18) patients (14 procedures), including but not limited to: <ul style="list-style-type: none"> • <u>Lead dislocation</u>: 28% (5/18) • <u>Lead breakage</u>: 28% (5/18) • <u>Revision or relocation due to pain at the pocket site</u>: 11% (2/18) • <u>Battery end of life</u>: 11% (2/18) • <u>Infection (“severe complication”)</u>: 0% (0/18) • <u>Neurological deficit (“severe complication”)</u>: 0% (0/18) • <u>Pain at IPG site</u>: 22% (4/18) | <p>when followed in the long-term (mean follow-up: 5.8 years), 50% of patients had at least one hardware complication requiring re-operation. Further, 22% of patients had pain at the IPG site. There were no cases of infection or neurological deficit.</p> |
| Hollingworth (2011) Cost-effectiveness study | KQ4 | <p><u>Population:</u> FBSS patients in the published Turner 2010 prospective cohort study of Washington State workers’ compensation patients*</p> <p><u>Diagnosis:</u> FBSS</p> <p><u>Intervention:</u> SCS (n = 51)</p> <p><u>Comparators:</u></p> <ol style="list-style-type: none"> 1. Pain clinic (PC) (n = 39) 2. Usual care (UC) (n = 68) <p><u>Cost-effectiveness analysis:</u></p> <ul style="list-style-type: none"> • Costs converted into 2007 US dollars • After first year of | <p>Recap of effectiveness results[†]:</p> <ul style="list-style-type: none"> • Effectiveness results used for this analysis were included in the original HTA[†] • <u>Primary outcome</u> (24 months): <ul style="list-style-type: none"> • SCS: 5% • PC: 3% • UC: 10% • No significant differences between any groups. <p>Cost-effectiveness results:</p> <ul style="list-style-type: none"> • Incremental cost per success (i.e., achieving the primary outcome): <u>SCS (n = 43) vs. UC (n = 61) (patients who completed 24 month f/u for primary outcome):</u> <ul style="list-style-type: none"> • Unadjusted incremental cost per patient achieving success on primary outcome: UC less costly, more effective (\$632,067: UC dominates) • Adjusted[‡] incremental cost per patient achieving success on primary outcome: \$334,704 (95% credible intervals, \$142,203 - \$489,243) • <u>SCS (n = 43) vs. PC (n = 34) (patients who completed 24 month f/u for primary outcome):</u> <ul style="list-style-type: none"> • Unadjusted incremental cost per patient achieving success on primary outcome:: \$846,977 | <ul style="list-style-type: none"> • In a cohort of Washington State workers’ compensation patients with FBSS, SCS was not cost-effective compared with usual care or pain clinic treatment. |

| Author (Year) Study type | Key Question | Demographics | Results | Conclusion |
|---|--------------|--|--|---|
| | | enrollment, costs discounted 3% <ul style="list-style-type: none"> Incremental cost-effectiveness defined as cost per successful outcome (i.e., additional cost of SCS/additional percentage of SCS patients achieving the primary outcome at 24 months) Primary outcome: composite of ≥ 50% leg pain relief relative to baseline, a 2-point or greater improvement in the Roland-Morris Disability Questionnaire, and less than daily opioid medication use | <ul style="list-style-type: none"> Adjusted[‡] incremental cost per patient achieving success on primary outcome: \$131,146 (95% credible intervals: \$271,075) (SCS dominates) <u>Permanent SCS implantation (n = 27) vs. PC (n = 22) (patients who completed 24 month f/u for cost data and primary outcome):</u> <ul style="list-style-type: none"> Unadjusted incremental cost per patient achieving success on primary outcome:: \$520,315 (PC dominates) Unadjusted incremental cost per patient achieving success on leg pain outcome: \$436,512 (PC dominates) Unadjusted incremental cost per patient achieving success on Roland Morris Disability Score: \$140,049 (PC dominates) | |
| Kemler (2010) Re-analysis of the data used in the NHS/NICE cost-effectiveness analysis of CRPS patients\$ (update not published by | KQ4 | <u>Population:</u> CRPS I patients as included in the Kemler 2000 RCT** (patient-level data available here but not in NHS analysis) <u>Diagnosis:</u> CRPS I <u>Intervention:</u> SCS + CMM | Cost-effectiveness results (SCS + CMM versus CMM): <ul style="list-style-type: none"> Cost difference: £6,994 higher with SCS QALY difference: 1.96 higher with SCS ICER (SCS relative to CMM): £3562 per QALY | <ul style="list-style-type: none"> SCS + CMM is cost-effective compared to CMM alone, with an 87% probability that SCS is cost-effective at a willingness to pay threshold of £30,000. The ICER of £3562 per QALY for SCS compared with CMM was lower than that reported in the NHS/NICE report (£25,095 per QALY) (and included in the original HTA) |

| Author (Year) Study type | Key Question | Demographics | Results | Conclusion |
|--|--------------|---|---------|------------|
| NHS/NICE Study sponsored by Medtronic | | <u>Comparator:</u> CMM alone <u>Cost-effectiveness analysis:</u> <ul style="list-style-type: none"> • NHS perspective • Time period: 15 years • Two-stage decision analytic tree and Markov model used • Primary outcome: \geq 50% pain relief at 6 months | | |

Abbreviations: CMM: conventional medical management; CRPS I: chronic regional pain syndrome I; FBSS: failed back surgery syndrome; F/U: follow-up; ICER: incremental cost-effectiveness ratio; IPG: implantable pulse generator; KQ: key question; NA: not applicable; NS: not statistically significant; RCT: randomized controlled trial; SCS: spinal cord stimulation; UC: usual care

* This study was included in the original HTA report to evaluate effectiveness of SCS versus pain clinic versus usual care. Reference: Turner JA, Hollingworth W, Comstock BA, Deyo RA. Spinal cord stimulation for failed back surgery syndrome: outcomes in a workers' compensation setting. *Pain* 2010;148:14-25.

† Conclusions from the original report:

Turner et al. (2010) “provided data on the short-term effectiveness of SCS compared with Pain Clinic and Usual Care treatments in FBSS patients with open workers’ compensation claims in the State of Washington. In general, the cohort study found no differences in outcomes between patients in the SCS and two control groups.

- **“Success” from a composite score:** There was no difference between SCS, pain clinic (PC), or usual care (UC) groups at any follow-up up to 24 months in the percent of patients achieving the primary outcome composite measure of success (includes pain, function, and medication usage components).
- **Pain relief:** Significantly more patients in the SCS group achieved \geq 50% leg pain relief by six months than those in the UC group, there was no difference between the SCS and PC group at the same follow-up; furthermore, no differences were identified between groups in the percentage of patients achieving leg pain relief of \geq 50% or more at the 12- and 24-month follow-ups.
- **Function:** There were no significant differences in either the Roland-Morris Disability Questionnaire (RDQ) scores or ability to perform daily tasks between treatment groups in the prospective cohort study on workers’ compensation patients.
- **Health-related quality of life (HR-QoL):** Reported no significant differences between treatment groups in SF-36 scores and work/disability status.

- **Medication usage:** Although significantly fewer patients in the SCS group used opioids on a less than daily basis than did those in the PC group at six months, no other significant differences between treatment groups were identified in the prospective cohort study on workers' compensation patients.

‡ Adjusted for baseline characteristics (cost in the year prior to enrollment, age, SF-36 mental health score, disability payments from another source, Roland-Morris Disability Questionnaire score, leg-pain intensity, duration of work time loss compensation, and legal representation)

§ NICE/NHS economic analysis included in original HTA; Simpson E, Duenas A, Holmes M, Papaioannou D, Chilcott J. Spinal cord stimulation for chronic pain of neuropathic or ischaemic origin: systematic review and economic evaluation. Health Technology Assessment 2009;13:1-179

** Kemler RCT included in original HTA; Kemler MA, Barendse GA, van Kleef M, et al. Spinal cord stimulation in patients with chronic reflex sympathetic dystrophy. N Engl J Med 2000;343:618-24.

2014 HTA Topics (New and Re-review): Response to Public Comments

March 24, 2014

Health Technology Assessment Program (HTA)

Washington State Health Care Authority

PO Box 42712

Olympia, WA 98504-2712

(360) 725-5126

www.hca.wa.gov/hta

shtap@hca.wa.gov

This document responds to all comments received on the Potential 2014 HTA Technology Topics. Public comment periods were accepted from March 1, 2014 through March 17, 2014. Comments were received from the following individuals and groups:

- John Medverd, MD, Washington State Radiological Society; Judith Jacobsen, MD, Washington Section Chair, and Judy Kimelman, MD, Washington Section Legislative Chair, The American Congress of Obstetricians and Gynecologists; Katie Kolan, JD, Director of Legislative and Regulatory Affairs, Washington State Medical Society.
- John Nadglowski, President and CEO, Obesity Action Coalition.
- Harrison Perry, Health Policy Analyst, American Academy of Otolaryngology- Head and Neck Surgery.
- Michael Bolen, Director, State Government Affairs, Medtronic, Inc.
- Ninh Nguyen, MD, FACS, FASMBS, President, American Society for Metabolic and Bariatric Surgery; Brian Sung, MD, FASMBS, Swedish Weight Loss Services; Judy Chen, MD, FASMBS, Swedish Weight Loss Services.

| Potential Topic | Comment | HCA Response |
|---|---|---|
| <p>John Medverd, MD and others representing multiple professional organizations</p> <p>Topic: Appropriate Breast Imaging for Breast Cancer Screening in Special Conditions</p> | <p>Complete comments with information attached below.</p> | <p>Thank you for providing information for the members of the Health Technology Clinical Committee for this proposed topics. All comments and attached information will be included in any future review.</p> <p>No change to proposed topics.</p> |
| <p>John Nadglowski, President and CEO, Obesity Action Coalition.</p> <p>Topic: Bariatric Surgery for Overweight/ Obese</p> | <p>Complete comments with information attached below.</p> | <p>Thank you for your comments. The information provided, including recent clinical practice guidelines, will be included in any future review. HCA has proposed review of bariatric surgery by the HTA program to ensure that care purchased by state programs is safe, effective and has value, especially as new populations are considered for this treatment.</p> <p>No change to proposed topics.</p> |
| <p>Harrison Peery, American Academy of Otolaryngology.</p> <p>Topic: Tympanostomy Tubes and Imaging for Rhinosinusitis</p> | <p>Complete comments with information attached below.</p> | <p>Thank you for your comments and for links to position statements and clinical guidelines for these topics. This information will be included in any future review of these topics.</p> <p>No change to proposed topics.</p> |
| <p>Michael Bolen, Director, State Government Affairs, Medtronic, Inc.</p> <p>Topic: Lumbar Spinal Fusion (re-review)</p> | <p>Complete comments with information attached below.</p> | <p>Thank you for providing comment and evidence for this proposed re-review. All information provided will be considered in any future re-review of lumbar spinal fusion.</p> <p>No change to proposed topics.</p> |
| <p>Ninh Nguyen, MD, FACS, FASMBS, President, American Society for Metabolic and Bariatric Surgery and others.</p> <p>Topic: Bariatric Surgery for Overweight/Obese</p> | <p>Complete comments with information attached below.</p> | <p>Thank you for your comments. The information provided by the American Society for Metabolic and Bariatric Surgery will be included in any future review.</p> <p>No change to proposed topics.</p> |

March 16, 2014

Dale Reisner, MD
President

Brian Seppi, MD
President-Elect

Nicholas Rajacich, MD
Past President

Ray Hsiao, MD
1st Vice President

Bruce Andison, MD
2nd Vice President

Shane Macaulay, MD
Secretary-Treasurer

Donna Smith, MD
Assistant Secretary-Treasurer

Jennifer Hanscom
Executive Director/CEO

Health Care Authority
Health Technology Assessment Program
626 8th Ave SE
P.O. Box 42712
Olympia, WA 98504-2712

Re: Appropriate Breast Imaging for Breast Cancer Screening in Special Conditions.

Dear Ms. Masters:

On behalf of the Washington State Medical Association and its over 9,800 physician and physician assistant members, we appreciate the opportunity to comment on the proposed topic for review.

The stated reasons for review and selection are concerns that include efficacy in special populations (e.g., high risk, dense tissue) and high state usage for breast imaging for breast cancer under special conditions. We offer the following information to the members of the Health Technology Clinical Committee (HTCC) for review and consideration.

Background: As background breast tissue density increases, the sensitivity of mammography is reduced. We also know that the recommendations for screening mammography are the same for women with dense breasts as for the rest of the population.

Mammography is the only screening modality that has undergone randomized controlled trials demonstrating a reduction in breast cancer mortality. **There is no recommendation that it be replaced with another test in any subset of the population.**

Additional screening options, generally:

Screening using MRI and ultrasound has been extensively tested. We know that breast density has more impact on the ability to detect cancer for mammography than it does for MRI or ultrasound. However, both ultrasound and MRI are associated with a much higher rate of benign biopsies. Moreover, both ultrasound and MRI have a higher rate of recommendation for short-interval follow-up than does mammography. **Therefore, when a patient chooses to have either an MRI or an ultrasound in addition to mammography, the patient should consider the benefit/risk trade-off of early cancer detection versus increased false positives.** This is where the relationship the patient has with her physician and the risk of the patient comes into play – the higher the patient's cancer risk, the more likely there will be benefit to additional testing.

Choosing to undergo additional screening is more favorable for women at high-risk than for average-risk women who simply have dense breasts.

Additional screening:

For patients who are interested in additional screening options, a breast cancer risk assessment may be appropriate. A breast cancer risk assessment can start the discussion between patient and physicians as to whether supplemental tests will be beneficial and what tests, if any, to order.

As noted above, there are additional breast “screening options” available. Those “screening options” include:

- MRI,
- Ultrasound, and
- Tomosynthesis (“3D mammography”).

MRI and patient risk: Screening using MRI has been shown to substantially increase the rate of cancer detection. Use of MRI screening is recommended for patients at very high-risk (>20% lifetime risk) and for those patients. Those patients with a personal history of breast cancer or a prior biopsy diagnosis of atypia (equivalent to a 15% to 20% lifetime risk) are considered at “intermediate risk” and in that case, a patient-centered shared decision-making approach is recommended.

Ultrasound and patient risk: Screening using ultrasound is often inaccessible to patients and is likely to incur an out-of-pocket cost to patients. Some studies have shown a high rate of false-positives but also a modest increase in breast cancer detection. A decision by a patient to use ultrasound should be made considering the risks, benefits and likely costs.

Tomosynthesis and patient risk: Screening using tomosynthesis (“3D mammography”) is sometimes being offered in addition to screening mammography but it has not yet been studied completely. The preliminary data is encouraging on the performance of tomosynthesis in women with dense tissue.

Cost of additional screening: While we have not performed an analysis of cost to Washington’s patients, California has reviewed topics related to the information this committee is reviewing. We offer for you the following excerpt from the California Breast Density Information Group (CBDIG)¹:

*The California legislature did not mandate insurance coverage for any supplemental breast cancer screening tests. Currently, **there are no insurance billing codes for screening breast ultrasound or tomosynthesis.** Screening breast MRI is usually covered for high risk women, but may not be for women at average risk who simply have dense breasts. As such, women who desire certain types of supplemental screening may be asked to pay out of pocket.*

¹ A working group of breast radiologists and breast cancer risk specialists, representing academic and community-based practices across California, formed to assist patients, referring doctors, and radiologists in responding to new legislation in California (SB 1538) that will mandate radiologists report breast density to patients. Source: <http://www.breastdensity.info/whoweare.html>, March 2013.

From a societal perspective, supplemental screening of the approximately 50 percent [those who under California's new mandatory additional law who would be recommended to seek additional screening] of California women with dense breasts would result in very substantial additional cost to the health care system.

Lastly, use of additional screening could divert needed resources to increase all women's access to screening mammography, the only screening modality which has been proven to reduce mortality in average risk women.

And, for your convenience, we have attached several supporting document for our review and consideration.

Sincerely,

Jonathan Medverd, MD
President
Washington State Radiological Society

Judith Jacobsen, MD
Washington Section Chair
American Congress of Obstetricians and Gynecologists

Judy Kimelman, MD
Washington Section Legislative Chair
American Congress of Obstetricians and Gynecologists

Katie Kolan, JD
Director of Legislative and Regulatory Affairs
Washington State Medical Association

CC:

Jennifer Hanscom, Executive Director/CEO, Washington State Medical Association
Executive Committee Members, Washington State Medical Association

Links to National Guidelines for Breast Cancer Screening:

American Cancer Society: <http://www.cancer.org/acs/groups/cid/documents/webcontent/003165-pdf.pdf>

American College of Radiology; Criteria and Diagnostics, 2012 :

<http://www.acr.org/~media/ACR/Documents/AppCriteria/Diagnostic/BreastCancerScreening.pdf>

References:

American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS) 5. Reston, Va: American College of Radiology; in press (anticipated 2013).

Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA : the journal of the American Medical Association* 2008;299:2151-63.

Berg WA, Zhang Z, Lehrer D, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA : the Journal of the American Medical Association* 2012;307:1394-404.

Bever TB, Anderson BO, Bonaccio E, et al. NCCN clinical practice guidelines in oncology: breast cancer screening and diagnosis. *Journal of the National Comprehensive Cancer Network : JNCCN* 2009;7(10):1060-96.

Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *The New England Journal of Medicine* 2007;356:227-36.

Buist DS, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40-49 years. *Journal of the National Cancer Institute* 2004;96:1432-40.

Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Annals of internal medicine* 2003;138:168-75.

Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. *Radiology* 2012;265:59-69.

Kelly KM, Dean J, Comulada WS, Lee SJ. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European radiology* 2010;20:734-42.

Kuhl C, Weigel S, Schrading S, et al. Prospective multicenter cohort study to refine management recommendations for women at elevated familial risk of breast cancer: the EVA trial. *Journal of*

clinical oncology : official journal of the American Society of Clinical Oncology 2010;28:1450-7.

Mainiero MB, Lourenco A, Mahoney MC, et al. ACR Appropriateness Criteria Breast Cancer Screening. Journal of the American College of Radiology : JACR 2013;10:11-4.

Pinsky RW, Helvie MA. Mammographic breast density: effect on imaging and breast cancer risk. Journal of the National Comprehensive Cancer Network : JNCCN 2010;8:1157-64; quiz 65.

Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA: a cancer journal for clinicians. 2007;57(2):75-89.

Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. Annals of internal medicine 2011;155:10-20.

Sickles EA. The use of breast imaging to screen women at high risk for cancer. Radiologic clinics of North America 2010; 48:859-878.

Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology 2013;267(1):47-56.

Smith RA, Brooks D, Cokkinides V, Saslow D, Brawley OW. Cancer screening in the United States, 2013. CA : A cancer journal for clinicians 2013; Epub ahead of print.

Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Annals of internal medicine 2008;148:337-47.

Vachon CM, van Gils CH, Sellers TA, et al. Mammographic density, breast cancer risk and risk prediction. Breast cancer research : BCR 2007;9:217.

van Gils CH, Otten JD, Verbeek AL, Hendriks JH. Mammographic breast density and risk of breast cancer: masking bias or causality? European journal of epidemiology 1998;14:315-20.

Whitehead J, Carlile T, Kopecky KJ, et al. Wolfe mammographic parenchymal patterns. A study of the masking hypothesis of Egan and Mosteller. Cancer 1985;56:1280-6.

Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncology. 2013;14(7):583-589.

Safety

The Use of Breast Imaging to Screen Women at High Risk for Cancer

Edward A. Sickles, MD

KEYWORDS

- Mammography • Breast MR imaging • Breast ultrasound
- Screening • Breast cancer • Women at high risk

Clinical experience and, more recently, genetic testing have identified a subset of women who are at substantially greater than average risk for developing breast cancer. The factors that define this group of women are diverse, ranging from previous biopsy showing cancer or high-risk lesions,¹⁻⁵ to strong family history of breast cancer,⁶ to previous exposure to high doses of chest radiation.⁷⁻¹¹ Several underlying genetic abnormalities have been discovered among some women at high very risk that add another level of complexity.¹²⁻²¹ Recent research has shown that breast density, as depicted at mammography, is a substantial independent risk factor for future breast cancer.²²⁻²⁷

Common to all women at high risk, by definition, is a greater likelihood of developing breast cancer. However, the cancers that occur in these women, especially those women with associated genetic abnormalities, tend to grow more rapidly,²⁸ tend to develop earlier in life,^{13,17} tend to be more difficult to identify at mammography,²⁹⁻³⁵ and may be less responsive to therapy.^{36,37} This quadruple whammy creates a sense of urgency that has caused most health care providers to adopt an intuitive, rather than rigorously scientific, approach in devising strategies to identify cancers in women at high risk. This article discusses a major part of this effort: the use of 3 breast imaging modalities (mammography, magnetic resonance [MR] imaging, and ultrasound [US] examination) to screen asymptomatic women at high risk for breast cancer.

FACTORS THAT PLACE WOMEN AT HIGH RISK FOR DEVELOPING BREAST CANCER

There are a wide variety of risk factors associated with the future development of breast cancer. Only those that play a substantial role in guiding clinical-care decisions regarding screening with breast imaging are considered here. High risk may be defined in terms of relative risk (the risk of individuals with a given risk factor divided by the risk of individuals without the same factor). As such, relative risks of 3 or more are often considered clinically relevant.³⁸ Other measures of risk are expressed as the percentage of women expected to develop cancer in a given future interval (per year, in the next given number of years, from current age until a given future age, lifetime risk).

Previous Breast Biopsy

Women who have already undergone biopsy for lesions diagnosed as malignant (ie, women with a personal history of breast cancer) are known to be at considerably higher than average risk for developing 1 or more additional cancers. The likelihood of contralateral breast cancer diagnosis among women with a personal history of breast cancer is 5% to 10% in the first decade after initial diagnosis, with some evidence that this risk is higher in younger than in older women.² A slightly higher level of risk (but for recurrence or second primary cancer in the ipsilateral breast), 5% to 10% over 5 years and 10% to 15% over 10 years, is reported for women undergoing breast conservation as

Department of Radiology, Box 1667, UCSF Medical Center, San Francisco, CA 94143-1667, USA
E-mail address: edward.sickles@ucsfmedctr.org

Radiol Clin N Am 48 (2010) 859–878

doi:10.1016/j.rcl.2010.06.012

0033-8389/10/\$ – see front matter © 2010 Elsevier Inc. All rights reserved.

6 of 77

treatment of a first cancer.³ Overall risk decreases in the second and subsequent decades after initial cancer diagnosis, at least partially because of decreased frequency of ipsilateral recurrence, but the risk during the first decade is high. However, all these risks are partially ameliorated if the patient is undergoing either systemic chemotherapy or chemoprevention.

Substantial risk also is imparted by a previous biopsy diagnosis (personal history) of epithelial ovarian cancer, especially for women younger than 50 years and within 10 years of the diagnosis of ovarian cancer.³⁹

Women who carry a biopsy diagnosis of lobular carcinoma in situ (LCIS) also are at considerably higher than average risk. Invasive cancers are more frequently lobular than ductal carcinoma, and are ipsilateral more frequently than contralateral to the LCIS, occurring at or near the site of LCIS diagnosis when ipsilateral.⁵ The rate of development of subsequent invasive carcinoma is estimated at 0.5% to 1% per year for LCIS, also reported as a 10% to 20% risk over 15 to 20 years and a 6- to 10-fold relative risk.⁴ Similar levels of high risk are reported for women who have a previous biopsy diagnosis of atypia. For both atypical lobular hyperplasia (ALH) and atypical ductal hyperplasia (ADH), in a study involving an average follow-up of 17 years, the relative risk of future invasive carcinoma is 4.2 for ALH and 4.3 for ADH, risks that are essentially doubled for women who also have a first-degree maternal relative with breast cancer (ALH 8.4, ADH 9.7).¹

Family History of Breast Cancer

There is an abundance of evidence demonstrating significant associations between a family history of breast cancer and subsequent development of breast cancer.⁶ As expected, these associations are strongest when first-degree relatives are involved. The future breast cancer risk to age 80 years of a 20-year-old woman with no family history is 7.8%, but the same woman has a risk of 13.3% with 1 first-degree relative and 21.1% with 2 first-degree relatives.⁶ Increased risk is imparted by either a maternal or paternal family history, although the infrequency of breast cancer among men causes most strong family history to involve mothers, sisters, and daughters. Among women who have a strong family history, risk is highest at young ages, and, for women of a given age, the risk is greater the younger the family member was when diagnosed with breast cancer.⁶ Despite the high relative risk at young ages, the absolute incidence of cancer is low, so the magnitude of this risk is tempered. For this

reason, most women with positive family history who develop breast cancer do so at either middle or old age. Substantial risk also is imparted by a family history of epithelial ovarian cancer.⁴⁰

Previous Mediastinal Radiation Therapy

Chest radiation therapy before age 30 years, primarily involving mantle irradiation for Hodgkin disease, is associated with an increased risk for developing breast cancer (latent period after treatment: mean 15–18 years, range 7–34 years).^{7,8,39} The frequency of subsequent breast cancer before age 40 years is reported to range from 31% to 50%.^{7,9,10,39} The relative risk of subsequent breast cancer diagnosis is highly dependent on age at irradiation, ranging from 136 for age at exposure less than 15 years, to 19 for ages 15 to 24 years, to 7 at ages 25 to 29 years, to 0.7 (no elevated risk) at ages 30 years and older.⁹ Risk also is directly proportional to radiation dose.³⁹

GENETIC ABNORMALITIES ASSOCIATED WITH A HIGH RISK OF DEVELOPING BREAST CANCER

Genetic abnormalities are believed to account for 5% to 10% of all cases of breast cancer.¹⁸ Women who are known or suspected carriers of *BRCA1* or *BRCA2* mutations are at especially high risk for developing breast (and ovarian) cancer, with an estimated breast cancer risk by age 70 years ranging from 46% to 87% for *BRCA1* mutation carriers (average 65%, 95% confidence interval 44%–78%) and 37% to 84% for *BRCA2* mutation carriers (average 45%, 95% confidence interval 31%–56%).^{12,13,17} Risks that are estimated from families with multiple cancer cases are at the high end of the observed range, approximately 85% risk for *BRCA1* and 65% for *BRCA2*.^{12,13,17} The risk imparted by both mutations is expressed at younger ages than in the general population.^{13,17} In one study involving *BRCA1* mutation carriers, the risk is reported as 3% by age 30 years, 19% by 40 years, and 51% by 50 years.¹² However, the relative risk associated with *BRCA1*, but not *BRCA2*, mutation declines significantly with advancing age,^{13,17} which likely accounts for the general observation that risk is expressed at somewhat younger ages for *BRCA1* than for *BRCA2* mutation carriers.¹⁷ The overall effect of this difference in relative risk is limited by the more substantial contribution imparted by increasing absolute incidence of breast cancer with advancing age, because the differences in relative risk occur primarily during years in which there is low absolute incidence.

There are also hereditary cancer predisposition syndromes associated with specific genetic abnormalities and an increased risk of breast cancer. Elevated risk is attributed to mutations of the *PTEN* gene in Cowden and Bannayan-Riley-Ruvalcaba syndrome, the *TP53* gene in Li-Fraumeni syndrome, the *MSH2* and *MLH1* genes in Muir-Torre syndrome, and the *STK11* gene in Peutz-Jeghers syndrome.^{14–16,18–21} The rarity of these syndromes precludes the calculation of reliable risk estimates for these syndromes.

Several sophisticated mathematical models have been developed to predict breast cancer risk based on family history and other factors. These models are used for risk assessment of individual patients, to decide whether to recommend genetic testing for *BRCA* gene mutations, to predict which patients may benefit from chemoprevention and/or high-risk screening, and by some investigators to determine study eligibility for research interventions designed for patients at high risk. Because these models are derived from different data sets, use different risk-calculating algorithms, and vary in the age to which they calculate cumulative breast cancer risk, each model indicates a somewhat different level of risk for the same woman. If more than 1 model is used to estimate risk for a given woman, it has been recommended to use the highest level of estimated risk.⁴¹

The Gail model, first and perhaps still the most widely used, is based on a woman's age, the number of first-degree relatives with breast cancer, age at menarche, age at first live birth, the number of previous breast biopsies (including presence/absence of atypia), and race/ethnicity (www.cancer.gov/bcrisktool/, accessed December 20, 2009).⁴² This is the only model validated for use among African American women as well as white women. However, this risk model omits consideration of age at diagnosis of first-degree relatives, any data on second-degree relatives, and paternal family history, limiting the value of the model as an estimator of risk associated with family history. The Claus model is based on more comprehensive family history data, including maternal and paternal family history, first- and second-degree relatives, as well as age at diagnosis of these relatives' cancers, albeit omitting family history data on ovarian cancer (www.palmgear.com/index.cfm?fuseaction=software.showsoftware&prodID=29820, accessed December 20, 2009).⁴³ Both of these models provide estimates of breast cancer risk, but do not specifically predict likelihood of the presence of *BRCA1* or *BRCA2* mutations. However, there are other models that do provide estimates for *BRCA* gene mutations. The BRCAPRO model

is based on both personal history and comprehensive family history data of breast and ovarian cancer (first- and second-degree relatives), as well as Ashkenazi Jewish ancestry (www4.utsouthwestern.edu/breasthealth/cagene/default.asp, accessed December 20, 2009).⁴⁴ The BOADICEA model incorporates data on family history of breast, ovarian, prostate, and pancreatic cancer, and also includes estimates of both breast and ovarian cancer risk (www.srl.cam.ac.uk/genepi/boadicea/boadicea_home.html, accessed December 20, 2009).⁴⁵ The Tyrer-Cuzick model uses not only comprehensive data on family history and Ashkenazi Jewish ancestry but also data on menstrual and reproductive history, previous biopsy showing either LCIS or atypia, height, and body mass index (www.ems-trials.org/riskevaluator, accessed December 20, 2009).⁴⁶

Although breast imaging facilities routinely collect most of the data required by the various models for calculating breast cancer and *BRCA* gene mutation risk, radiologists generally have neither training nor expertise in using the models, performing the calculations, and interpreting the results. Therefore, many find it helpful to partner with a nearby high-risk clinic, staffed by health care professionals who specialize in risk assessment and genetic counseling. However, for many women, the radiologists who interpret their breast imaging examinations may be more aware of the likelihood that they may be at high risk for future breast cancer than any of their other health care providers. For this reason, it is advisable for radiologists to develop a basic understanding of risk assessment, to know whether a given request for high-risk screening is appropriate.

MAMMOGRAPHIC BREAST DENSITY ASSOCIATED WITH A HIGH RISK FOR DEVELOPING BREAST CANCER

There is a growing body of evidence of an independent association between breast density as depicted at mammography and subsequent risk of developing breast cancer. Observational studies have shown a statistically significant increase in relative risk for progressively increasing categories of breast density.^{22–27} In the United States and in many other countries, breast density is usually described according to the 4 categories defined in the Breast Imaging Data and Reporting System (BI-RADS) of the American College of Radiology (ACR): almost entirely fatty (0%–24%), scattered areas of fibroglandular density (25%–50%), heterogeneously dense (51%–75%), and extremely dense (>75%).⁴⁷ The most frequently

described result is that the relative risk of women having extremely dense breasts ($\geq 75\%$ dense) is 4 to 6 times that of women with almost entirely fatty breasts (0%, $<1\%$, or $<10\%$ dense).^{22–27}

However, these seemingly high levels of risk may be misleading. All the breast cancer risk factors described earlier are present in only a small minority of women, so it is clinically relevant to consider absence of these risk factors as representing average risk. This is not the case for mammographic breast density. Fewer than 10% of screening mammography examinations in the United States are categorized during image interpretation as showing almost entirely fatty breasts, and fewer than 10% of examinations are interpreted as showing the breasts to be extremely dense.⁴⁸ Therefore, more than 80% of screening examinations show breasts with scattered areas of fibroglandular density or heterogeneously dense breasts. The relative risk for the heterogeneously dense compared with the scattered areas category is less than 1.5 in all studies.^{22–27} Compared with breasts of average density (those approximately 50% dense, at or near the threshold between the scattered areas and heterogeneously dense categories), the relative risk for heterogeneously dense breasts is less than 1.2 in all studies, and the relative risk for extremely dense breasts is less than 2.1.^{22–27}

BREAST IMAGING APPROACHES TO SCREENING ASYMPTOMATIC WOMEN AT HIGH RISK FOR CANCER

In devising strategies to screen asymptomatic women at high risk for breast cancer, the following types of evidence should be assessed, in order of decreasing quality and reliability: systematic reviews and meta-analyses of randomized controlled trials (RCTs), individual RCTs, non-randomized intervention studies (cohort and case-control studies), observational studies, nonexperimental studies (case series), and expert opinion.⁴⁹ The only high-quality evidence supporting recommendations for routine periodic cancer screening using breast imaging come from the several RCTs of screening mammography performed in the United States and Europe, involving nearly 500,000 subjects.^{50–57} All but 1 of these RCTs demonstrated a statistically significant reduction in breast cancer mortality among the populations invited to screening. Overall, based on a recent meta-analysis of the RCTs, there was a 20% reduction in breast cancer mortality among the women invited to screening.⁵⁸ However, none of the RCTs were designed to study screening mammography among women of any age at high

risk. Therefore, screening strategies for these women must be based on less-scientifically rigorous evidence. Given the clinical context that there is a known subset of women at very high risk (hence, sufficiently high disease prevalence and incidence), coupled with the combination of rapid tumor growth, early age at diagnosis, lower than usual sensitivity of screening mammography, and possible poor response to therapy, most health care providers are willing to accept more intuitive approaches to screening, based on observational studies, case series, and expert opinion.

Within the past few years, several national medical organizations have developed guidelines and recommendations for screening women at high risk using breast imaging. First was the American Cancer Society (ACS), which published guidelines for screening with MR imaging as an adjunct to mammography in 2007.⁵⁹ More recently, the National Comprehensive Cancer Network (NCCN) issued parallel guidelines,⁶⁰ followed by recommendations for mammography, MR imaging, and US made jointly by the Society of Breast Imaging (SBI) and the ACR.⁶¹ There are only minor differences in the guidelines and recommendations of these several organizations, although those of the SBI/ACR are more comprehensive. This article discusses separately the roles of mammography, MR imaging, and US in screening women at high risk, primarily because the strengths and limitations of these imaging modalities differ considerably, resulting in different recommendations for their use.

Mammography

As stated earlier, screening mammography is the only breast imaging modality validated by multiple RCTs and meta-analyses to reduce breast cancer mortality. Most national medical organizations in the United States, weighing this and other benefits against the several harms of screening (false-positive-induced anxiety, inconvenience, cost, and occasional morbidity, as well as overdiagnosis), have endorsed the routine annual use of mammography beginning at age 40 years. These organizations include the ACS, NCCN, SBI, ACR, the American College of Obstetricians and Gynecologists (every 1–2 years at ages 40–49 years), and the American College of Surgeons. The widespread use of screening mammography in the past several decades has contributed substantially to the nearly 30% reduction in breast cancer mortality observed in the United States.^{56,62,63}

Given that routine screening mammography is already recommended in almost all (if not all)

countries that have widely available mammographic service (annually, starting at age 40 years in the United States), the primary screening mammography issue for women at high risk is whether to lower the age at which screening should begin, and if so, to what age.

There are theoretic advantages and limitations to screening mammography in women less than the age of 40 years. Potential advantages include the long life expectancy of younger women and the low frequency of comorbidity. Reduction in breast cancer mortality in younger women, in the absence of substantial comorbidity, by definition would add many quality-adjusted life years (QALY), the most frequently used metric in assessing cost-effectiveness. If one is willing to infer the existence of some mortality reduction, by extrapolation from RCT data involving older women, as well as the observation of similar cancer detection rates in screening women at high risk aged 30 to 39 years versus women at average risk aged 40 to 49 years (Table 1), then there would be substantial benefit.

Potential limitations to screening mammography in women less than the age of 40 years include the reduced frequency of breast cancer, somewhat reduced sensitivity of mammography, slightly increased radiation risk, and the suggestion of an increase in recall rate. However, the recall rate issue seems to be spurious, because the apparent overall increase in recall rate among women less than 40 years of age disappears when examinations are segregated by availability of previous examinations for comparison; absence of previous examinations is the strongest factor producing increased recall rate,⁶⁴ and younger women are more likely to have baseline examinations (Table 2).

The risk of radiation oncogenesis imparted at mammography is exceptionally small, because of the very low doses imparted at mammography and the relative insensitivity of the mature breast to the carcinogenic effects of ionizing radiation. At the usual lower age limit for screening (age 40

years), estimates of radiation risk are tiny, and are considered minor in comparison with the proven benefits of breast cancer mortality reduction.⁶⁵ However, at younger ages, especially less than 30 years, the breast seems to be more radio-sensitive, not only as demonstrated by the age-dependent risk among women with Hodgkin disease who received chest radiation therapy before the age of 30 years,⁹ but also among Japanese atomic bomb survivors, women receiving radiation therapy for postpartum mastitis, fibroadenomatosis, and other benign breast conditions, as well as women exposed to high cumulative doses of chest radiation when undergoing repeated fluoroscopic monitoring of artificial pneumothorax as treatment of tuberculosis.⁶⁵ In the age range 30 to 39 years, for which benefit is presumed but not proven and for which radiation risk is only slightly higher than at age 40 years and older, expert opinion indicates that benefit substantially exceeds risk, so concerns about radiation oncogenesis should not limit the acceptability of screening mammography in women at high risk for future breast cancer.⁶¹ Benefit/risk ratios are somewhat less favorable for women in their 20s, but, among this age cohort, it is observed that more than half of the benefit and less than half of the radiation risk is found in women aged 25 to 29 years, so expert opinion supports the acceptability of screening mammography down to age 25 years, but only for those women at the highest levels of future breast cancer risk.⁶¹

The sensitivity of screening mammography has also been shown to be age-dependent, lower in younger than in older women, and lowest for women younger than 40 years.^{66,67} The major explanation for this reduced sensitivity is the tumor-masking effect of dense fibroglandular tissue, which is present more frequently in younger women. For the subset of women at very high risk younger than 40 years who have also undergone screening MR imaging, the sensitivity of mammography is very low indeed, principally because of the

Table 1
Cancer detection rates for screening mammography at ages 30 to 39 years for women at high risk and at ages 40 to 49 years for all women

| Age Range (y) | Examinations | Cancers | Cancer Detection Rate (/1000) |
|---------------|--------------|---------|-------------------------------|
| 30–39 | 3252 | 11 | 3.4 |
| 40–49 | 65,209 | 216 | 3.3 |

Data are from the screening mammography practice at the University of California San Francisco (UCSF) Medical Center, from 1985 to 2009. At UCSF, screening mammography is recommended at ages 30 to 39 years only for women at high risk.

Table 2

Recall rates for screening mammography at less than 40 years of age for women at high risk and at ages 40 to 70 years and older than 70 years for all women, depending on whether interpreted in comparison with previous examination(s)

| Age Range (y) | Examinations | Recalls | Recall Rate (%) |
|---|--------------|---------|-----------------|
| All examinations | | | |
| <40 | 4727 | 336 | 7.11 |
| 40–70 | 108,087 | 6245 | 5.78 |
| >70 | 27,730 | 1266 | 4.57 |
| Total | 140,544 | 7847 | 5.58 |
| Examinations interpreted in comparison with previous examination(s) | | | |
| <40 | 1450 | 53 | 3.66 |
| 40–70 | 86,489 | 3784 | 4.38 |
| >70 | 24,426 | 918 | 3.76 |
| Total | 112,365 | 4755 | 4.23 |
| Examinations interpreted without comparison with previous examination(s) | | | |
| <40 | 3277 | 283 | 8.64 |
| 40–70 | 21,598 | 2461 | 11.39 |
| >70 | 3304 | 348 | 10.53 |
| Total | 28,179 | 3092 | 10.97 |

Data are from the screening mammography practice at the UCSF Medical Center, from 1997 to 2009. At UCSF, screening mammography is recommended at less than 40 years of age only for women at high risk. Recalls = all abnormal interpretations (BI-RADS assessment categories 0, 4, or 5). Recall rate = number of recalls divided by number of examinations. When considering all examinations, the recall rate seems to be lower for women less than 40 years of age than for older women, but, after accounting for whether examinations were interpreted in comparison with previous examination(s), this difference in recall rate is no longer present. A likely explanation is that a greater proportion of women less than 40 years of age are undergoing baseline examinations, at which there are no previous examinations and at which the recall rate is known to be significantly and substantially higher. This explanation, based on less extensive data, initially was proposed in Sickles EA. Successful methods to reduce false-positive mammography interpretations. *Radiol Clin North Am* 2000;38(4):693–700.

many mammographically and clinically occult cancers that are identified at MR imaging.⁴¹ However, despite lower sensitivity (all women <40 years of age) or very much lower sensitivity (women <40 years of age who also undergo screening MR imaging), there still is presumed benefit in screening with mammography, because some mammographically visible cancers (especially ductal carcinoma in situ [DCIS] presenting as a small group of microcalcifications) are not depicted at MR imaging.⁴¹ Furthermore, the ongoing switch among radiology practices from the use of screen film to digital mammography is likely to result in improved sensitivity in screening women at high risk younger than 40 years, because the use of digital mammography has been shown to result in superior performance (including higher sensitivity) for women aged less than 50 years who are pre- or perimenopausal and who have dense breasts.^{68,69}

This leaves the observed low frequency of breast cancer as the principal limitation to

screening women less than 40 years of age. In the United States, fewer than 7% of all breast cancers are found in women younger than 40 years.⁷⁰ Described in other terms, the likelihood of breast cancer diagnosis in the next 10 years for an 30-year-old woman at average risk is only 0.4% (1 in 250).³⁸ The efficacy of screening is reduced if there is a low prior probability of disease in the screened population, and this is a major reason why screening mammography is not recommended for women at average risk in their 30s. However, given the proven efficacy and widely accepted use of screening mammography for women at average risk beginning at 40 years of age, it seems reasonable to screen women at high risk at any age less than 40 years if their current risk is either equal to or greater than that of a 40-year-old woman at average risk, because these women have a sufficiently high prior probability of disease.

Ultimately, the decision about when to begin high-risk screening mammography depends on

the balance between benefits and harms. In terms of benefit, women with the full spectrum of breast cancer risk discussed earlier (except for the lower risk levels imparted by dense breasts alone) can be expected to have risk levels in their 30s exceeding that of a 40-year-old woman at average risk, so it is reasonable to support high-risk screening mammography in this decade. Because breast cancer is exceedingly rare among women less than the 30 years of age,⁷⁰ except for *BRCA* mutation carriers and in other exceptional circumstances, age 30 years seems to be the appropriate lower age limit for screening mammography among women at high risk. For known carriers of the *BRCA1* mutation, breast cancer risk is sufficiently high at age 20 years to possibly justify screening,⁴¹ but concerns about radiation risk among such young women suggest that, instead, the beginning age for screening mammography should not be before age 25 years.⁶¹ The breast cancer risk of *BRCA2* mutation carriers also is sufficiently high to support the start of screening mammography at age 25 years.⁴¹

Among the guidelines and recommendations recently issued by national medical organizations in the United States, those made jointly by the SBI and ACR specifically list the indications for

screening mammography in women at high risk. The ACS and NCCN guidelines discuss screening mammography only by inference, in that they involve screening MR imaging as an adjunct to mammography (hence the assumption that mammography is also recommended). The ACS and NCCN did not consider the indications for screening mammography in women at high risk other than those who should also undergo screening MR imaging. The more comprehensive SBI/ACR recommendations are listed in **Table 3**, with footnotes indicating the inferred guidelines of the ACS and NCCN.

MR Imaging

RCTs, cohort studies, and case-control studies have not been performed to assess the efficacy of screening MR imaging to reduce breast cancer mortality. Therefore, the efficacy of screening MR imaging must be estimated based on less robust data, and it must be remembered that, at best, such data provide inferential evidence rather than scientific proof.

There are several prospective observational studies of screening MR imaging as an adjunct to mammography. This article discusses those

Table 3
Recommendations on periodic high-risk screening mammography made by national medical organizations in the United States

| Indication | Age to Start (y) |
|---|-----------------------|
| <i>BRCA</i> mutation carriers ^{a,b} | 25–30 ^c |
| Untested first-degree relatives of <i>BRCA</i> mutation carriers ^{a,b} | 25–30 ^c |
| Lifetime breast cancer risk 20% or greater ^{a,b,d} | Variable ^e |
| Chest radiation therapy between age 10 and 30 y ^{a,b} | Variable ^f |
| Personal history of breast cancer (invasive carcinoma, DCIS) | At diagnosis |
| Previous breast biopsy showing LCIS ^b | At diagnosis |
| Previous breast biopsy showing ADH or ALH | At diagnosis |
| Personal history of invasive ovarian carcinoma | At diagnosis |
| Mother or sister with early-onset breast cancer | Variable ^e |

These are the recommendations of the SBI and ACR.⁶¹ The recommendations involve annual screening and continue to the age at which routine screening mammography is recommended for women at average risk, which, in the United States, is 40 years according to these and many other national medical organizations.

^a Recommendations attributed to the ACS, by inference, in that they involve screening MR imaging as an adjunct to mammography.⁵⁹ These recommendations include women with Li-Fraumeni, Cowden, and Bannayan-Riley-Ruvalcaba syndromes and their first-degree relatives (not recommended by SBI/ACR).

^b Recommendations attributed to the NCCN, by inference, in that they involve screening MR imaging as an adjunct to mammography.⁶⁰ These recommendations include women with Li-Fraumeni, Cowden, and Bannayan-Riley-Ruvalcaba syndromes and their first-degree relatives (not recommended by SBI/ACR).

^c Age 30 years but not before age 25 years.

^d As defined by a risk-assessment model that is largely dependent on family history (Gail model excluded).

^e Age 30 years (but not before age 25 years), or 10 years earlier than age at diagnosis of youngest affected relative, whichever is later.

^f Eight years after radiation therapy, but not before age 25 years.

studies involving nonoverlapping patient populations in which multiple rounds of screening were performed and in which at least 10 cancers were detected at screening. In chronologic sequence by date of publication, these include the Dutch multi-institution study reported by Kriege and colleagues,²⁹ the single-institution Canadian study reported by Warner and colleagues,³⁰ the United Kingdom multi-institution study reported by Leach and colleagues,³¹ the single-institution German study reported by Kuhl and colleagues,³² the multi-institution Italian study reported by Sardaneli and colleagues,³³ the multi-institution Norwegian study reported by Hagen and colleagues,³⁴ and the Austrian single-institution study reported by Riedl and colleagues.³⁵

Interstudy comparisons are of limited value because the eligibility criteria for entry into these various studies differ substantially, as do other important aspects of study design, including that some, but not all, of the studies also involve supplemental screening with US. However, because all the studies involve patient populations heavily weighted with women at high risk, it is reasonable to analyze study findings in combination as well as individually (Table 4). When doing so, several observations are clear. The sensitivity of screening MR imaging is substantially higher than that of either mammography or US (or the combination of mammography and US); almost all detectable cancers are identified at screening with the combination of mammography and MR imaging; most of the malignancies detected only at MR imaging are small, node-negative, invasive cancers; and screening MR imaging results in a low interval cancer rate (see Table 4).

The superior ability of screening MR imaging to detect clinically occult breast cancer is especially apparent in the subset of women at very high risk involving carriers of *BRCA* mutations,⁷¹ in whom screening with mammography alone results in limited success.^{29,72-74} In this select patient population, the disparity in sensitivity between MR imaging and mammography is greater, and the cancers detected at MR imaging are small, frequently node-negative, with few interval cancers.⁷¹

The sensitivity of mammography combined with MR imaging permits detection of virtually all detectable cancers. As shown in Table 4, the sensitivity of screening MR imaging across all studies is 81%, which is increased to 92% by its use in combination with screening mammography. Given the interval cancer rate of approximately 6% across all studies, the incremental benefit of further adding screening US seems to be minimal. In the American College of Radiology Imaging

Network (ACRIN) high-risk screening study, supplemental screening MR imaging was provided to 627 self-selected women shortly after completion of the third and final screening round involving mammography and US.⁷⁵ The addition of prevalence MR imaging screening doubled the number of cancers detected in this incidence screening round (from 8 to 16); only 2 of the 16 cancers would have been missed had screening been limited to mammography and MR imaging.⁷⁵ It is now widely accepted that, if screening includes both mammography and MR imaging in a given woman at high risk, there is no need for additional US screening.⁴¹

In clinical practice, some breast imaging centers perform screening mammography and MR imaging concurrently, whereas others prefer to stagger the examinations every 6 months (for example, mammography in October and MR imaging in April). The theoretic advantage of staggered examinations is the potential for detection 6 months earlier of cancers that would be visible at both examinations; this amounts to approximately 25% to 30% of detectable cancers (see Table 4). However, there are also practical advantages to concurrent screening; in addition to increased patient convenience, it may be helpful to compare a borderline abnormal MR imaging examination with concurrently performed mammography to establish the benignity of the MR imaging finding(s). Given the absence of data comparing the efficacies of concurrent and staggered approaches, either is considered acceptable. In my practice, we provide staggered screening to our patients at high risk, except for those for whom examinations on 2 separate days would be inconvenient.

Insofar as the success of supplemental screening MR imaging is dependent on limitations to the success of screening mammography among women at high risk, one would expect superior MR imaging performance among women with denser breasts. However, in a population of *BRCA* mutation carriers screened with both modalities, the sensitivities of MR imaging versus mammography are reported as 86% versus 18% for women with primarily dense (>50% dense) breasts, and 94% versus 33% for women with primarily fatty (\leq 50% dense) breasts.⁷⁶ Therefore, supplemental MR imaging screening seems to have benefit that is independent of mammographic breast density.

As discussed earlier, the decision about whether to recommend high-risk screening depends on the balance between benefits and harms. For MR imaging, the observed benefits of increased sensitivity, detection of small node-negative cancers,

and a low interval cancer rate seem to be substantial, although these permit only the inference of reduced breast cancer mortality. The observed harms of screening MR imaging include false-positive interpretations (resulting in recall for additional breast imaging work-up but not cancer diagnosis), a high percentage of probably benign assessments (resulting in short-interval follow-up MR imaging examinations that rarely lead to cancer diagnosis), and recommendations for biopsy (that do not produce a cancer diagnosis). **Table 4** displays the frequencies of false-positive results for screening MR imaging, as best can be determined from the source articles (which do not describe false-positive outcomes in as much detail as true-positive outcomes). All but 1 of the studies of screening MR imaging also reports at least 1 calculation of specificity, albeit with variable definitions of false-positive examinations. Specificity is uniformly reported to be lower for screening MR imaging than for screening mammography, indicating poorer performance.^{29–33,35} Furthermore, the overall frequencies of probably benign assessments (8%) and biopsies performed for MR imaging–detected abnormalities (4%), shown in **Table 4**, are considerably higher than those observed for screening mammography among women at average risk.⁷⁷

However, these unfavorable false-positive results are balanced, at least somewhat, by the additional report that the likelihood of malignancy (positive predictive value) among lesions undergoing biopsy based on screening MR imaging is acceptably high (41%, see **Table 4**), being somewhat higher than that reported for screening mammography in women at average risk (25%).⁷⁷ Such a higher positive predictive value in screening women at high risk might be expected, because of the greater prior probability of cancer.

The other factor that limits the effect of the high false-positives reported for screening MR imaging is that false-positive MR imaging rates are reported to be substantially lower at incidence screening than at the initial prevalence screen.³⁰ This likely occurs not only because of the learning process for screening MR imaging interpretation such that increased experience leads to fewer false-positives (as already shown for mammography),⁷⁸ but also because false-positive rates generally are lower when current examinations are compared with previous examinations.⁶⁴

Most health care providers seem willing to consider that the inferred benefits of screening MR imaging (substantial numbers of additional cancers detected) outweigh the observed harms of high false-positives. This is mirrored by the same national medical organizations that

recommend screening mammography for women at high risk also recommending supplemental screening with MR imaging.^{59–61} The ACS, the first organization to develop screening MR imaging guidelines, recommends screening MR imaging only for women at very high risk (several specific risk factors, as well as lifetime risk estimated to be at least 20% using risk assessment models that are largely dependent on family history). The ACS recommends against screening MR imaging for women at a lifetime risk of less than 15% (the woman at average risk has a lifetime risk of approximately 12% when estimated at age 20 years). The ACS states that current evidence is insufficient to recommend either for or against screening MR imaging for women at high risk at a lifetime risk between 15% and 20%, with further clarification that screening decisions should be made on a case-by-case basis and that payment should not be a barrier.⁵⁹ The joint SBI/ACR recommendations and the NCCN guidelines are similar to those of the ACS (details are given in **Table 5**).

The guidelines and recommendations of all these national organizations use estimates of lifetime risk, which decreases with advancing age despite breast cancer incidence progressively increasing. As a result, the lifetime risk of a 60-year-old woman, for example, may be substantially lower than her risk at age 30 years, whereas her risk of breast cancer diagnosis in the next several years may be considerably higher at age 60 years than at 30 years. The net effect is that a woman's lifetime risk inaccurately estimates the (perhaps more clinically relevant) shorter-term risk that she and her health care provider often consider in deciding on screening MR imaging. As a result, some have proposed that a shorter term than lifetime risk may provide a more appropriate basis for screening guidelines.⁴¹ In this regard, a 10-year interval of risk may be the most useful, because this time period is sufficiently long to indicate substantial benefit (assuming the presence of benefit), but it is also sufficiently short to overcome the limitations inherent in lifetime estimates of risk. Based on data derived from the Claus risk assessment model, Berg⁴¹ asserts that the use of a 5% 10-year-risk threshold for recommending screening MR imaging identifies groups of women similar to those who have a lifetime risk estimated to be at least 20%.

Despite the ACS suggestion that payment should not be a barrier in deciding on screening MR imaging, the examination indeed is expensive in comparison with mammography and US, especially when annual examinations are

Table 4
Prospective studies of screening MR imaging involving multiple screening rounds, in which at least 10 cancers were identified

| Country | Women/Screening Examinations | Cancers ^a | Invasive Cancers | Node Positive ^b | Cancers Detected at Mammography ^c |
|---------------------|------------------------------|----------------------|------------------|----------------------------|--|
| Netherlands | 1909/4169 | 45 | 39 | 5 (13) | 18 (40) |
| Canada | 236/457 | 22 | 16 | 2 (13) | 8 (36) |
| United Kingdom | 649/1881 | 35 | 29 | 5 (17) | 14 (40) |
| Germany | 529/1452 | 43 | 34 | 5 (15) | 14 (33) |
| Italy ^h | 278/377 | 15 | 11 | 1 (9) | 9 (60) |
| Norway ⁱ | 491/867 | 21 | 18 | 6 (26) | 9 (43) |
| Austria | 327/672 | 27 | 16 | 2 (13) | 13 (48) |
| Total | 4419/9875 | 208 | 163 | 26 (16) | 85 (41) |

Cancers Detected at:

| Country | US ^c | MR Imaging ^c | Mammography + US ^c | Mammography + MR Imaging ^c | Interval Cancers ^c |
|---------------------|-----------------|-------------------------|-------------------------------|---------------------------------------|-------------------------------|
| Netherlands | — | 32 (71) | — | 40 (89) | 4 (9) |
| Canada | 7 (32) | 17 (77) | 12 (55) | 20 (91) | 1 (5) |
| United Kingdom | — | 27 (77) | — | 33 (94) | 2 (6) |
| Germany | 17 (40) | 39 (91) | 21 (49) | 40 (93) | 1 (2) |
| Italy ^h | 9 (60) | 13 (87) | 10 (67) | 15 (100) | 0 |
| Norway ⁱ | — | 17 (81) | — | 18 (86) | 2 (10) |
| Austria | 11 (41) | 23 (85) | 13 (48) | 25 (93) | 2 (7) |
| Total | 44 (41) | 168 (81) | 56 (52) | 191 (92) | 12 (6) |

| Country | MR Imaging False-positives | | | PPV ₃ (Biopsy Performed) ^g |
|---------------------|----------------------------|------------------------------|-------------------------------|--|
| | Recalls ^d | Probably Benign ^e | Biopsy Performed ^f | |
| Netherlands | 177 (4) | 275 (7) | 56 (1) | 32 (57) |
| Canada | 58 (13) | 24 (5) | 32 (7) | 17 (53) |
| United Kingdom | 202 (11) | 169 (9) | 77 (4) | 27 (35) |
| Germany | — | 167 (12) | 78 (5) | 39 (50) |
| Italy ^h | — | — | 24 (6) | 13 (54) |
| Norway ⁱ | — | — | — | — |
| Austria | — | — | 101 (15) | 23 (23) |
| Total | 437 (7) | 635 (8) | 368 (4) | 151 (41) |

^a Breast cancers only, including both screen-detected and interval cancers. All studies report data for only the most advanced cancer diagnosed (1 cancer per woman).

^b Number of node-positive invasive cancers (%).

^c Number of cancers (% of all cancers). The percentage calculation represents sensitivity.

^d Number of MR imaging examinations assessed as BI-RADS category 0, 4, or 5 (% of examinations).

^e Number of MR imaging examinations assessed as BI-RADS category 3 (% of examinations).

^f Number of MR imaging examinations assessed as BI-RADS category 0, 3, 4, or 5 for which fine-needle aspiration, core, or surgical biopsy subsequently was performed (% of examinations).

^g PPV, positive predictive value. Number of cancers detected at MR imaging (number of cancers detected at MR imaging divided by the number of biopsies performed for MR imaging-detected abnormalities, ^f expressed as %).

^h Three of 18 cancers are excluded because either mammography, US, or MR imaging was not performed.

ⁱ Four of 25 cancers are excluded because either mammography or MR imaging was not performed.

Table 5
Recommendations on periodic high-risk screening MR imaging made by national medical organizations in the United States

| Indication | Age to Start (y) |
|--|------------------|
| <i>BRCA</i> mutation carriers | 30 |
| Untested first-degree relatives of <i>BRCA</i> mutation carriers | 30 |
| Lifetime breast cancer risk 20% or greater | 30 |
| Chest radiation therapy between age 10 and 30 y | Variable |

These are the recommendations of the SBI and ACR.⁶¹ The recommendations involve annual screening, and continue to the age at which routine screening mammography stops because of limited life expectancy, substantial comorbidity, or unwillingness to undergo additional testing (including biopsy), if recommended, and cancer treatment, if appropriate. The ACS and NCCN also recommend screening MR imaging for women with Li-Fraumeni, Cowden, and Bannayan-Riley-Ruvalcaba syndromes and their first-degree relatives, starting at age 30 years. The NCCN also recommends screening MR imaging for women with a biopsy diagnosis of LCIS (starting at diagnosis), based on 1 single-institution study that demonstrates cancer detection in 5 of 135 women.

Data from Port ER, Park A, Borgen PI, et al. Results of MR imaging screening for breast cancer in high-risk patients with LCIS and atypical hyperplasia. *Ann Surg Oncol* 2007;14(3):1051–7.

planned. A recent cost-effectiveness study reported acceptable cost per QALY-gained data (<\$100,000) for annual screening MR imaging, but only for women at highest risk (known *BRCA1* mutation carriers from 35–59 years of age and known *BRCA2* mutation carriers from 40–49 years of age).⁷⁹ This same study reports the cost-effectiveness in QALY gained for screening mammography for the even wider age range of 25 to 69 years as being highly favorable, only \$18,952 and \$28,421 for *BRCA1* and *BRCA2* carriers, respectively; the study also reports that the incremental effect of adding screening MR imaging from ages 25 to 29 years to any range of older women adds more than \$300,000 per QALY gained.⁷⁹

Comparison of the data in **Tables 3** and **5** shows a large discrepancy in the groups of women recommended for screening mammography and MR imaging, according to SBI/ACR guidelines. The basic difference is that women in the intermediate-risk group (several specific risk factors and 15%–20% lifetime risk) are recommended for screening mammography at ages 30 to 39 years, but not routinely for screening MR imaging at any age (only on a case-by-case basis). There are several reasons to recommend screening mammography among these women: inference of a benefit in mortality reduction from extrapolation of RCT results, mammography for women in this group yields a similar cancer detection rate to that for women at average risk in their 40s (see **Table 1**), and women in this group have a breast cancer risk equal to or greater than that of a 40-year-old woman at average risk. Screening MR imaging is not routinely recommended in these

women, at least in part because, compared with mammography, it has more frequent false-positives, is more expensive, takes longer to perform and interpret, requires an intravenous injection, and has more frequent intolerances (reports from 2 high-risk screening studies indicate MR imaging nonparticipation rates of 9% and 42%).^{80,81}

US

As with screening MR imaging, RCTs, cohort studies, and case-control studies have not been completed to assess the efficacy of screening US to reduce breast cancer mortality. Therefore, the efficacy of screening US also must be estimated based on less robust data, and, at best, such data provide inferential evidence rather than scientific proof.

However, there are several single-institution observational studies of screening US as an adjunct to mammography. The eligibility criteria for entry into these studies differ substantially, as do other important aspects of study design. Nonetheless, because the studies each involve patient populations heavily weighted with women at high risk, it is reasonable to analyze study findings in combination. Berg³⁸ has summarized the outcomes reported in these studies, comprising almost 50,000 examinations (see **Table 2** in Ref.³⁸). Overall, the incremental cancer detection rate provided by screening US is 3.6 per 1000 examinations, 94% of the cancers are invasive, more than 70% are 1 cm in size or smaller, and 86% are node-negative.³⁸ However, the potential effect of the encouraging results reported in these several studies is

limited by several aspects of experimental design: performance of each study at only a single institution or breast imaging practice, US interpretive criteria used in each study not fully described, only 1 round of screening US performed (1 of the studies included multiple rounds of screening but did not separately report outcomes at prevalence and incidence screening), interpretation of screening US examinations not done independently of mammography, false-positive outcomes reported incompletely, and interval cancer rates not reported at all.

To overcome these deficiencies, as well as to provide more complete evidence on the efficacy of screening US, ACRIN has conducted a multi-institution prospective study involving 3 rounds of screening.⁸² The first screening round resulted in incremental cancer detection outcomes that were strikingly similar to those reported for the several previous single-institution studies, albeit with high false-positive rates.⁸³ Preliminary results from the 2 incidence screening rounds were presented recently, showing similar cancer detection rates, but somewhat lower false-positive rates than those observed at prevalence screening.^{84,85}

There are particularly pertinent data from the 4 screening MR imaging studies among women at high risk that also include screening US (discussed earlier), 1 of which is a multi-institution study. Outcomes are summarized in **Table 4**. Overall, these studies show modest incremental rates of cancer detection and sensitivity for screening US beyond what is detected at mammography, similar to the results reported for the several single-institution US-only studies and for the multi-institution ACRIN study. However, these studies also show that screening US performance in detecting cancer is far inferior to that of screening MR imaging (at false-positive rates similar to those of MR imaging), and that screening US provides no substantial incremental increase in cancer detection beyond that achieved by the combination of screening mammography and MR imaging (see **Table 4**). Therefore, screening MR imaging seems to be more effective than screening US as an adjunctive examination to mammography for women at very high risk, leading to the suggestion that screening US should be considered only if MR imaging is unavailable, impractical, or poorly tolerated by a given woman.

Although MR imaging is the preferred examination to screen women at very high risk, some have proposed the use of screening US for those women in the so-called intermediate-risk group of women (those with a personal history of breast

cancer, previous biopsy diagnosis of LCIS or atypia, or a family history from which lifetime risk of breast cancer is estimated at 15%–20%) who, after consideration on a case-by-case basis, do not undergo screening MR imaging.⁴¹ Despite its limited efficacy compared with MR imaging, there are several reasons why screening US might be considered in this scenario. It is more readily available, less expensive, better tolerated, and does not require intravenous injection. However, when used as in almost all the reported observational studies, as a physician-performed examination using a hand-held transducer, screening US is a more time-consuming examination for the interpreting physician than MR imaging, one that may become a major time sink. In the ACRIN study, the median duration of a physician-performed screening US examination was 19 minutes.⁸³ For this reason, many breast imaging practices have decided not to provide screening US services for this, or any other, indication.

Even more problematic when considering workforce issues is the consideration of whether screening US should be offered to all women with mammographically dense breasts. There already is evidence among women at intermediate risk demonstrating incremental cancer detection beyond that achieved by screening mammography, albeit at much higher false-positive rates.^{83,86–92} However, as discussed earlier, the presence of dense breasts alone may not impart enough risk to justify screening (because the added risk beyond that of women with average-density breasts is not particularly high). For women with dense breasts with no other substantial risk factors, who are at only slightly higher than average risk, why not simply consider the use of screening US for purposes of incremental cancer detection?⁶¹ Given the extremely large number of women who might be screened (approximately half of women undergoing screening mammography have heterogeneously or extremely dense breasts, increasing to 90% if women with scattered areas of fibroglandular density also are included),⁴⁸ this might be reasonable if the magnitude of incremental cancer detection is similar to that demonstrated for intermediate risk and women at very high risk (not yet known), if false-positive rates are acceptably low (not yet known), if the cost of examination is acceptably low (similar to that of screening mammography), and if workforce issues are solved by automated rather than physician-performed hand-held approaches to screening US (encouraging results are already reported for one automated US screening device).⁹³ Among women with primarily fatty breasts, screening US as an adjunct to mammography

has been shown to be ineffective in incremental cancer detection.^{38,89–91} So, until sufficient evidence is reported to justify the use of screening US for all women with dense breasts, perhaps the better way to consider this issue is that, for now, fatty breasts are a contraindication to screening US.

Two large-scale studies involving screening US are already underway that may provide the necessary evidence on the usefulness of screening US for women with dense breasts. An RCT has been started in Japan, designed to study 50,000 women with screening mammography and handheld US performed by a technologist or a physician and then interpreted by a physician (and 50,000 controls with screening mammography only).⁹⁴ The defined study population is women aged 40 to 49 years, because this is the age range in Japan at which breast cancer incidence peaks, and because a high percentage of Japanese women in this age range have dense breasts. The primary end points of this trial are sensitivity and specificity, so data on both incremental cancer detection and false-positives should be forthcoming. The rate of advanced cancers will also be measured, because this has been demonstrated in the screening mammography RCTs to be a surrogate for reduction in breast cancer mortality.⁵⁸ However, this trial has several limitations: the screening interval is 2 years, despite evidence that screening mammography at age 40 to 49 years is more effective with annual screening^{58,95}; the study population being so different from those in Western countries may limit the generalization of study outcomes; and the study likely is underpowered to provide follow-up data on breast cancer deaths because of the low breast cancer risk of native Japanese women, and also because women with fatty breasts are not excluded from the study.

The second study is a nonrandomized multi-institution effort involving multiple annual screening rounds, conducted primarily in the United States, using a matched-pair design similar to that of the ACRIN study, assessing the performance of screening mammography alone versus the combination of screening mammography and US. However, this is a much larger-scale study than the ACRIN study (approximately 25,000 women with dense breasts), with no emphasis on recruiting an especially high-risk population of women with dense breasts, using automated rather than hand-held US, and involving US scanning of only the dense portions of the breasts.

Interpretation of both mammography and US examinations is done using the standard batch-reading approach used for screening examinations, with mammography and US examinations read independently by different radiologists. The primary study end point is sensitivity; the rate of advanced cancers is also measured. Study strengths include use of a more workforce-efficient, purely screening approach to US, and large-scale multi-institution design in a population of women likely to be representative of all women with dense breasts in Western countries. Study limitations are non-randomized experimental design and inability to study breast cancer deaths as an end point.

One other issue concerning screening US deserves discussion. Almost all current data comparing the performance of mammography alone with that of mammography plus US involves screen-film mammography.^{30,32,33,86–92} However, digital mammography is currently in the process of replacing screen-film mammography, largely because of research that shows superior performance of digital mammography for women younger than 50 years, who are pre- or perimenopausal, and who have dense breasts.^{68,69} As indicated earlier, the improved performance of adjunctive screening with US is observed primarily in women with dense breasts,^{38,89–91} most of whom also are younger than 50 years and pre- or perimenopausal. Therefore, it is likely that the gradual but steady switch from screen film to digital mammography is resulting in improved mammography performance and, consequently, in diminished benefit for supplemental screening with US. However, the potentially limiting effect of digital mammography on screening US performance, not incorporated into most of the data produced by the already reported studies, is indeed factored into the currently ongoing hand-held US study in Japan and the automated US study in the United States, because the mammography component in both these studies involves digital mammography alone.

Other Imaging Modalities

Breast imaging modalities other than mammography, MR imaging, and US have also been used for purposes of cancer detection. Nuclear medicine imaging, initially using standard gamma cameras,^{96,97} but more recently with higher-resolution breast-specific equipment,⁹⁸ and also with positron emission scanning,⁹⁹ has been performed principally in the diagnostic setting, although there is 1 preliminary report

suggesting possible efficacy for screening women at high risk.¹⁰⁰ Breast-specific computed tomography (CT) scanning, initially studied in the late 1970s using prototype equipment that then was abandoned,^{101,102} is now being investigated with renewed interest. Several research groups are developing dedicated breast CT scanners, and a preliminary report indicates some degree of incremental cancer detection beyond that provided by mammography.¹⁰³ Most promising in this regard is the use of contrast-enhanced CT scanning, which, theoretically, may provide similar cancer detection ability to that currently achieved by screening MR imaging.¹⁰⁴ Thermography, although widely evaluated in the 1970s and found to be ineffective for screening,^{105,106} continues to be modified and, presumably, improved; the most recent adaptations of thermographic imaging have not been tested extensively. Breast transillumination,^{107–109} electrical impedance spectroscopy,^{110,111} near infrared spectroscopy,^{110,112} and microwave imaging^{110,113} also have yet to demonstrate screening efficacy. In summary, current screening applications involving any of these modalities are considered investigational.

Future Developments

Active research is underway involving new applications of mammography, MR imaging, and US technology. For mammography, several tomography approaches are being developed that promise to increase the detection of some cancers currently not depicted in dense breasts using either screen-film or digital imaging, simultaneously substantially reducing the frequency of false-positive screening outcomes by verifying the benignity of most recalled lesions that are determined to represent summation artifact (superimposition of normal breast structures) at diagnostic mammography.^{114–116} For MR imaging, new applications using either diffusion imaging^{117,118} or spectroscopy^{119–122} may permit early cancer detection without the need for intravenous contrast injection. For both MR imaging and US, studies of the differential elasticity of benign and malignant breast structures may permit earlier detection of some cancers that currently are not depicted.^{123–125} If 1 or more of these currently investigational applications eventually demonstrates screening efficacy, especially efficacy of considerable magnitude, the current balance of usefulness among mammography, MR imaging, and US may change, perhaps substantially.

SUMMARY

Although there currently is no evidence of reduced breast cancer mortality for screening women at high risk with mammography, MR imaging, or US, the presumptive evidence of early cancer detection provided by numerous observational studies has led to the publication of guidelines and recommendations for the selective use of these imaging modalities. In general, annual screening mammography is recommended for women of appropriately high risk beginning at age 30 years, supplemental screening with MR imaging is recommended for a subset of women at very high risk, and screening US is suggested for women for whom MR imaging is appropriate but unavailable, impractical, or poorly tolerated. The use of screening US remains controversial among women who have no substantial risk factors other than dense breasts.

REFERENCES

1. Page DL, Dupont WD, Rogers LW, et al. Atypical hyperplastic lesions of the female breast: a long-term follow-up study. *Cancer* 1985;55(11):2698–708.
2. Fowble B, Hanlon A, Freedman G, et al. Second cancers after conservative surgery and radiation for stages I-II breast cancer: identifying a subset of women at increased risk. *Int J Radiat Oncol Biol Phys* 2001;51(3):679–90.
3. Dershaw DD. Breast imaging and the conservative treatment of breast cancer. *Radiol Clin North Am* 2002;40(3):501–16.
4. Arpino G, Laucirica R, Elledge RM. Premalignant and in situ breast disease: biology and clinical implications. *Ann Intern Med* 2005;143(6):446–57.
5. Li CI, Malone KE, Saltzman BS, et al. Risk of invasive breast carcinoma among women diagnosed with ductal carcinoma in situ and lobular carcinoma in situ. *Cancer* 2006;106(10):2104–12.
6. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 2001;358(9291):1389–99.
7. Dershaw DD, Yahalom J, Petrek JA. Breast carcinoma in women previously treated for Hodgkin disease: mammographic evaluation. *Radiology* 1992;184(2):421–3.
8. Yahalom J, Petrek JA, Biddinger PW, et al. Breast cancer in patients irradiated for Hodgkin's disease: a clinical and pathological analysis of 45 events in 37 patients. *J Clin Oncol* 1992;10(11):1674–81.

9. Hancock SL, Tucker MA, Hoppe RT. Breast cancer after treatment of Hodgkin's disease. *J Natl Cancer Inst* 1993;85(1):25–31.
10. Bhatia S, Robison LL, Oberlin O, et al. Breast cancer and other second neoplasms after childhood Hodgkin's disease. *N Engl J Med* 1996;334(12):745–51.
11. Travis LB, Hill DA, Dores GM, et al. Breast cancer following radiotherapy and chemotherapy among young women with Hodgkin disease. *JAMA* 2003;290(4):465–75.
12. Easton DF, Ford D, Bishop DT, et al. Breast and ovarian cancer incidence in BRCA1-mutation carriers. *Am J Hum Genet* 1995;56(1):265–71.
13. Ford D, Easton DF. The genetics of breast and ovarian cancer. *Br J Cancer* 1995;72(4):805–12.
14. Propeck PA, Warner T, Scanlan KA. Sebaceous carcinoma of the breast in a patient with Muir-Torre syndrome. *AJR Am J Roentgenol* 2000;174(2):541–2.
15. Chen J, Lindblom A. Germline mutation screening of the STK11/LKB1 gene in familial breast cancer with LOH on 19p. *Clin Genet* 2000;57(5):394–7.
16. Srivastava A, McKinnon W, Wood ME. Risk of breast and ovarian cancer in women with strong family histories. *Oncology (Williston Park)* 2001;15(7):889–902.
17. Antoniou A, Pharoah PD, Narod S, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* 2003;72(5):1117–30.
18. Garber JE, Offit K. Hereditary cancer predisposition syndromes. *J Clin Oncol* 2005;23(2):276–92.
19. Newman BD, Maher JF, Subauste JS, et al. Clustering of sebaceous gland carcinoma, papillary thyroid carcinoma and breast cancer in a woman as a new cancer susceptibility disorder: a case report. *J Med Case Reports* 2009;3:6905.
20. Hobert JA, Eng C. PTEN hamartoma tumor syndrome: an overview. *Genet Med* 2009;11(10):687–94.
21. Clements A, Robison K, Granai C, et al. A case of Peutz-Jeghers syndrome with breast cancer, bilateral sex cord tumor with annular tubules, and adenoma malignum caused by STK11 gene mutation. *Int J Gynecol Cancer* 2009;19(9):1591–4.
22. Boyd NF, Byng JW, Jong A, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *J Natl Cancer Inst* 1995;87(9):670–5.
23. Byrne C, Schairer C, Wolfe J, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *J Natl Cancer Inst* 1995;87:1622–9.
24. Ursin G, Ma H, Wu AH, et al. Mammographic density and breast cancer in three ethnic groups. *Cancer Epidemiol Biomarkers Prev* 2003;12(4):332–8.
25. Vacek PM, Geller BM. A prospective study of breast cancer risk using routine mammographic breast density measurements. *Cancer Epidemiol Biomarkers Prev* 2004;13(5):715–22.
26. Boyd NF, Martin LJ, Sun L, et al. Body size, mammographic density, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2006;15(11):2086–92.
27. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356(3):227–36.
28. Tilanus-Linthorst MM, Kriege M, Boetes C, et al. Hereditary breast cancer growth rates and its impact on screening policy. *Eur J Cancer* 2005;41(11):1610–7.
29. Kriege M, Brekelmans CTM, Boetes C, et al. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *N Engl J Med* 2004;351(5):427–37.
30. Warner E, Plewes DB, Hill KA, et al. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA* 2004;292(11):1317–25.
31. Leach MO, Boggis CRM, Dixon AK, et al. Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). *Lancet* 2005;365(9473):1769–78.
32. Kuhl CK, Schrading S, Leutner CC, et al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *J Clin Oncol* 2005;23(33):8469–76.
33. Sardanelli F, Podo F, D'Agnolo G, et al. Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRIT study): interim results. *Radiology* 2007;242(3):698–715.
34. Hagen AI, Kvistad KA, Maehle L, et al. Sensitivity of MRI versus conventional screening in the diagnosis of BRCA-associated breast cancer in a national prospective series. *Breast* 2007;16(4):367–74.
35. Riedl CC, Pehold L, Flöry D, et al. Magnetic resonance imaging of the breast improves detection of invasive cancer, preinvasive cancer, and premalignant lesions during surveillance of women at high risk for breast cancer. *Clin Cancer Res* 2007;13(20):6144–52.
36. Robson M, Levin D, Federici M, et al. Breast conservation therapy for invasive breast cancer in

- Ashkenazi women with BRCA gene founder mutations. *J Natl Cancer Inst* 1999;91(24):2112–7.
37. Foulkes WD, Chappuis PO, Wong N, et al. Primary node negative breast cancer in BRCA1 mutation carriers has a poor outcome. *Ann Oncol* 2000; 11(3):307–13.
 38. Berg WA. Beyond standard mammographic screening: mammography at age extremes, ultrasound, and MR imaging. *Radiol Clin North Am* 2007;45(5):895–906.
 39. Travis LB, Curtis RE, Boice JD Jr, et al. Second malignant neoplasms among long-term survivors of ovarian cancer. *Cancer Res* 1996;56(7):1564–70.
 40. Lee JS, John EM, McGuire V, et al. Breast and ovarian cancer in relatives of cancer patients, with and without BRCA mutations. *Cancer Epidemiol Biomarkers Prev* 2006;15(2):359–63.
 41. Berg WA. Tailored supplemental screening for breast cancer: what now and what next? *AJR Am J Roentgenol* 2009;192(2):390–9.
 42. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989; 81(24):1879–86.
 43. Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer. *Cancer* 1994;73(3):643–51.
 44. Berry DA, Iversen ES Jr, Gudbjartsson DF, et al. BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes. *J Clin Oncol* 2002; 20(11):2701–12.
 45. Antoniou AC, Cunningham AP, Peto J, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer* 2008;98(8):1457–66.
 46. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* 2004;23(7):1111–30.
 47. D'Orsi CJ, Bassett LW, Berg WA, et al. Breast imaging reporting and data system – mammography: ACR BI-RADS® – mammography. 4th edition. Reston (VA): American College of Radiology; 2003. p. 179–89.
 48. D'Orsi CJ, Sickles EA, Bassett LW, et al. Breast imaging reporting and data system – mammography: ACR BI-RADS® – mammography. 5th edition. Reston (VA): American College of Radiology, in press.
 49. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *Br Med J* 2001;323(7308):334–6.
 50. Shapiro S, Venet W, Strax P, et al. Periodic screening for breast cancer: the health insurance plan project and its sequelae, 1963–1986. Baltimore (MD): Johns Hopkins University Press; 1988.
 51. Alexander FE, Anderson TJ, Brown HK, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* 1999; 353(9168):1903–8.
 52. Tabár L, Vitak B, Chen HH, et al. The Swedish two-county trial twenty years later: updated mortality results and new insights from long-term follow-up. *Radiol Clin North Am* 2000;38(4):625–51.
 53. Miller AB, To T, Baines CJ, et al. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years. *J Natl Cancer Inst* 2000;92(18):1490–9.
 54. Nyström L, Andersson I, Bjurstam N, et al. Long-term effects of mammography screening: updated overview of the Swedish randomized trials. *Lancet* 2002;359(9310):909–19.
 55. Miller AB, To T, Baines CJ. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. *Ann Intern Med* 2002;137(5 Part 1):305–12.
 56. Otto SJ, Fracheboud J, Looman CWN, et al. Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. *Lancet* 2003; 361(9367):1411–7.
 57. Bjurstam N, Bjorneld L, Warwick J, et al. The Gothenburg breast screening trial. *Cancer* 2003; 97(10):2387–96.
 58. Smith RA, Duffy SW, Gabe R, et al. The randomized trials of breast cancer screening: what have we learned? *Radiol Clin North Am* 2004;42(5):793–806.
 59. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007;57(2):75–89.
 60. Lehman CD, Smith RA. The role of MRI in breast cancer screening. *J Natl Compr Canc Netw* 2009; 7(10):1109–15.
 61. Lee CH, Dershaw DD, Kopans D, et al. Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *J Am Coll Radiol* 2010;7(1):18–27.
 62. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2009. *CA Cancer J Clin* 2009;59(4):225–49.
 63. Age-adjusted United States death rates by year (female breast - invasive cancer). SEER cancer statistics review 1975–2006. Available at: seer.cancer.gov/csr/1975_2006/results_single/sect_04_table.08.pdf. Accessed December 20, 2009.
 64. Sickles EA. Successful methods to reduce false-positive mammography interpretations. *Radiol Clin North Am* 2000;38(4):693–700.
 65. Rothenberg LN, Feig SA, Haus AG, et al. A guide to mammography and other breast imaging

- procedures: NCRP report no. 149. Bethesda (MD): National Council on Radiation Protection and Measurements; 2004. p. 252–66.
66. Kerlikowske K, Grady D, Barclay J, et al. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *JAMA* 1996;276(1):33–8.
 67. Rosenberg RD, Hunt WC, Williamson MR, et al. Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183,134 screening mammograms in Albuquerque, New Mexico. *Radiology* 1998;209(2):511–8.
 68. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 2005;353(17):1773–83.
 69. Pisano ED, Hendrick RE, Yaffe MJ, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. *Radiology* 2008;246(2):376–83.
 70. Age distribution of incidence cancers (female breast - invasive). SEER cancer statistics review 1975–2006. Available at: seer.cancer.gov/csr/1975_2006/results_single/sect_01_table.10_2pgs.pdf. Accessed December 20, 2009.
 71. Causer PA, Jong RA, Warner E, et al. Breast cancers detected with imaging screening in the *BRCA* population: emphasis on MR imaging with histopathologic correlation. *Radiographics* 2007;27(Suppl 1):S165–82.
 72. Brekelmans CTM, Seynaeve C, Bartels CCM, et al. Effectiveness of breast cancer surveillance in *BRCA1/2* gene mutation carriers and women with high familial risk. *J Clin Oncol* 2001;19(4):924–30.
 73. Møller P, Borg Å, Evans DG, et al. Survival in prospectively ascertained familial breast cancer: analysis of a series stratified by tumour characteristics, *BRCA* mutations and oophorectomy. *Int J Cancer* 2002;101(6):555–9.
 74. Komenaka IK, Dittkoff B-A, Joseph K-A, et al. The development of interval breast malignancies in patients with *BRCA* mutations. *Cancer* 2004;100(10):2079–83.
 75. Berg WA, Zhang Z, Cormack JB, et al. Supplemental yield and performance characteristics of screening MRI after combined ultrasound and mammography: ACRIN 6666 American College of Radiology Imaging Network. In: Programs and abstracts of the 95th Scientific Assembly and Annual Meeting of the Radiological Society of North America. Chicago, November 29 to December 4, 2009. p. 103.
 76. Bigenwald RZ, Warner E, Gunasekara A, et al. Is mammography adequate for screening women with inherited *BRCA* mutations and low breast density? *Cancer Epidemiol Biomarkers Prev* 2008;17(3):706–11.
 77. Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. *Radiology* 2006;241(1):55–66.
 78. Miglioretti DL, Gard CC, Carney PA, et al. When radiologists perform best: the learning curve in screening mammogram interpretation. *Radiology* 2009;253(3):632–40.
 79. Plevritis SK, Kurian AW, Sigal BM, et al. Cost-effectiveness of screening *BRCA1/2* mutation carriers with breast magnetic resonance imaging. *JAMA* 2006;295(20):2374–84.
 80. Sardanelli F, Podo F, Santoro F, et al. Compliance for mammography, ultrasound, and contrast-enhanced MRI during a multicenter surveillance of women at high risk for breast cancer. In: Programs and abstracts of the 95th Scientific Assembly and Annual Meeting of the Radiological Society of North America. Chicago, November 29 to December 4, 2009. p. 102.
 81. Berg WA, Blume JD, Adams AM, et al. Reasons women at elevated risk of breast cancer refuse breast MR imaging screening: ACRIN 6666. *Radiology* 2010;254(1):79–87.
 82. Berg WA. Rationale for a trial of screening breast ultrasound: American College of Radiology Imaging Network (ACRIN) 6666. *AJR Am J Roentgenol* 2003;180(5):1225–8.
 83. Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA* 2008;299(18):2151–63.
 84. Berg WA, Zhang Z, Cormack JB, et al. Screening breast ultrasound as a supplement to mammography: yield of annual screening in ACRIN 6666 American College of Radiology Imaging Network. In: Programs and abstracts of the 95th Scientific Assembly and Annual Meeting of the Radiological Society of North America. Chicago, November 29 to December 4, 2009. p. 100–1.
 85. Berg WA, Zhang Z, Marques H, et al. False positives induced by annual screening US added to mammography: ACRIN 6666 American College of Radiology Imaging Network. In: Programs and abstracts of the 95th Scientific Assembly and Annual Meeting of the Radiological Society of North America. Chicago, November 29 to December 4, 2009. p. 102.
 86. Gordon PB, Goldenberg SL. Malignant breast masses detected only by ultrasound: a retrospective review. *Cancer* 1995;76(4):626–30.

87. Buchberger W, Niehoff A, Obrist P, et al. Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. *Semin Ultrasound CT MR* 2000;21(4):325–36.
88. Kaplan SS. Clinical utility of bilateral whole-breast US in the evaluation of women with dense breast tissue. *Radiology* 2001;221(3):641–9.
89. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology* 2002;225(1):165–75.
90. Leconte I, Feger C, Galant C, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. *AJR Am J Roentgenol* 2003;180(6):1675–9.
91. Crystal P, Strano SD, Shcharynski S, et al. Using sonography to screen women with mammographically dense breasts. *AJR Am J Roentgenol* 2003;181(1):177–82.
92. Corsetti V, Ferrari A, Ghirardi M, et al. Role of ultrasonography in detecting mammographically occult breast carcinoma in women with dense breasts. *Radiol Med* 2006;111(3):440–8.
93. Kelly KM, Dean J, Comulada WS, et al. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *Eur Radiol* 2010;20(3):734–42.
94. Ohuchi N, Suzuki A, Sakurai Y, et al. Current status and problems of breast cancer screening. *Japan Medical Association Journal* 2009;52(1):45–9.
95. Duffy SW, Day NE, Tabár L, et al. Markov models for breast tumor progression: estimates from empirical screening data and implications for screening. In: *Programs and abstracts of the NIH Consensus Development Conference on Breast Cancer Screening for Women Ages 40–49*. Bethesda (MD); 1997. p. 91–2.
96. Khalkhali I, Villanueva-Meyer J, Edell SL, et al. Diagnostic accuracy of ^{99m}Tc-sestamibi breast imaging: multicenter trial results. *J Nucl Med* 2000;41(12):1973–9.
97. Khalkhali I, Baum JK, Villanueva-Meyer J, et al. ^{99m}Tc-sestamibi breast imaging for the examination of patients with dense and fatty breasts: multicenter study. *Radiology* 2002;222(1):149–55.
98. Brem RF, Floerke AC, Rapelyea JA, et al. Breast-specific gamma imaging as an adjunct imaging modality for the diagnosis of breast cancer. *Radiology* 2008;247(3):651–7.
99. Berg WA, Weinberg IN, Narayanan D, et al. High-resolution fluorodeoxyglucose positron emission tomography with compression (“positron emission mammography”) is highly accurate in depicting primary breast cancer. *Breast J* 2006;12(4):309–23.
100. Brem RF, Rapelyea JA, Zisman G, et al. Occult breast cancer: scintimammography with high-resolution breast-specific gamma camera in women at high risk for breast cancer. *Radiology* 2005;237(1):274–80.
101. Chang CHJ, Sibala JL, Fritz SL, et al. Specific value of computed tomographic breast scanner (CT/M) in diagnosis of breast disease. *Radiology* 1979;132(3):647–52.
102. Gisvold JJ, Reese DF, Karsell PR. Computed tomographic mammography (CTM). *AJR Am J Roentgenol* 1979;133(6):1143–9.
103. Lindfors KK, Boone JM, Nelson TR, et al. Dedicated breast CT: initial clinical experience. *Radiology* 2008;246(3):725–33.
104. Lindfors KK. Breast CT. In: *Programs and abstracts of the 9th Postgraduate Course of the Society of Breast Imaging*. Colorado Springs (CO), 2009.
105. Moskowitz M, Milbrath J, Gartside P, et al. Lack of efficacy of thermography as a screening tool for minimal and stage I breast cancer. *N Engl J Med* 1976;295(5):249–52.
106. Moskowitz M. Screening for breast cancer: how effective are our tests? A critical review. *CA Cancer J Clin* 1983;33(1):26–39.
107. Gisvold JJ, Brown LR, Swee RG, et al. Comparison of mammography and transillumination light scanning in the detection of breast lesions. *AJR Am J Roentgenol* 1986;147(1):191–4.
108. Monsees B, Destouet JM, Totty WG. Light scanning versus mammography in breast cancer detection. *Radiology* 1987;163(2):463–5.
109. Monsees B, Destouet JM, Gersell D. Light scan evaluation of nonpalpable breast lesions. *Radiology* 1987;163(2):467–70.
110. Poplack SP, Paulsen KD, Hartov A, et al. Electromagnetic breast imaging: average tissue property values in women with negative clinical findings. *Radiology* 2004;231(2):571–80.
111. Zheng B, Zuley ML, Sumkin JH, et al. Detection of breast abnormalities using a prototype resonance electrical impedance spectroscopy system: a preliminary study. *Med Phys* 2008;35(7):3041–8.
112. Pogue BW, Poplack SP, McBride TO, et al. Quantitative hemoglobin tomography with diffuse near-infrared spectroscopy: pilot results in the breast. *Radiology* 2001;218(1):261–6.
113. Meaney PM, Fanning MW, Reynolds T, et al. Initial clinical experience with microwave breast imaging in women with normal mammography. *Acad Radiol* 2007;14(2):207–18.
114. Poplack SP, Tosteson TD, Kogel CA, et al. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening

- mammography. *AJR Am J Roentgenol* 2007; 189(3):616–23.
115. Rafferty EA. Digital mammography: novel applications. *Radiol Clin North Am* 2007;45(5):831–43.
 116. Gur D, Abrams GS, Chough DM, et al. Digital breast tomosynthesis: observer performance study. *AJR Am J Roentgenol* 2009;193(2): 586–91.
 117. Yoshikawa MI, Ohsumi S, Sugata S, et al. Comparison of breast cancer detection by diffusion-weighted magnetic resonance imaging and mammography. *Radiat Med* 2007;25(5): 218–23.
 118. Park MJ, Cha ES, Kang BJ, et al. The role of diffusion-weighted imaging and the apparent diffusion coefficient (ADC) values for breast tumors. *Korean J Radiol* 2007;8(5):390–6.
 119. Bartella L, Morris EA, Dershaw DD, et al. Proton MR spectroscopy with choline peak as malignancy marker improves positive predictive value for breast cancer diagnosis: preliminary study. *Radiology* 2006;239(3):686–92.
 120. Bartella L, Thakur SB, Morris EA, et al. Enhancing nonmass lesions in the breast: evaluation with proton (^1H) MR spectroscopy. *Radiology* 2007; 245(1):80–7.
 121. Sardanelli F, Fausto A, Di Leo G, et al. In vivo proton spectroscopy of the breast using the total choline peak integral as a marker of malignancy. *AJR Am J Roentgenol* 2009;192(6):1608–17.
 122. Tozaki M, Fukuma E. ^1H MR spectroscopy and diffusion-weighted imaging of the breast: are they useful tools for characterizing breast lesions before biopsy? *AJR Am J Roentgenol* 2009; 193(3):840–9.
 123. McKnight AL, Kugel JL, Rossman PJ, et al. MR elastography of breast cancer: preliminary results. *AJR Am J Roentgenol* 2002;178(6):1411–7.
 124. Itoh A, Ueno E, Tohno E, et al. Breast disease: clinical application of US elastography for diagnosis. *Radiology* 2006;239(2):341–50.
 125. Ginat DT, Destounis SV, Barr RG, et al. US elastography of breast and prostate lesions. *Radiographics* 2009;29(7):2007–16.

The Breast Density Conundrum¹

Carl J. D'Orsi, MD

The article in this month's issue of *Radiology* by the California Breast Density Information Group (CBDIG) (1) is a balanced evaluation of the clinical impact of breast density both as a risk factor for the development of breast cancer and in reducing sensitivity for early detection of breast cancer. To briefly summarize the CBDIG review of the literature, at the clinical (individual patient) level, the cancer risk imparted by breast density is less important than the masking effect of dense breast tissue on the ability of mammography to depict early cancer. This conclusion is not unique to the CBDIG; it is repeated in the new edition of the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) (2), forming part of the rationale for eliminating quartile estimates of percentage dense tissue when classifying breast density into four categories for use in the mammography report. By using only the text descriptions for the two denser categories (heterogeneously and extremely dense), which specifically relate to the masking effect of dense breast tissue, the interpreting radiologist will classify breast density in a more clinically relevant manner.

Complicating discussion of the density-related effects of cancer risk and masking is legislation in nine states (including California) mandating the reporting of breast density in the lay letter sent to all women undergoing mammography, accompanied by text that suggests possible benefit if supplementary screening is added to the mammographic examination already performed. Because these are two distinctly separate issues with more robust data supporting the masking effect of breast density, the legislation, which tends to blend these issues, may cause confusion for women sent information regarding their "dense" breast status. As

with most complex issues, there is not a "one-size-fits-all" answer.

The CBDIG article also describes the relative strengths and limitations of supplementary screening with either magnetic resonance (MR) imaging or ultrasonography (US) for women with dense breasts, appropriately recommending a nuanced approach based primarily on the likelihood of malignancy for a given woman as indicated by her own specific breast cancer risk factors. For the very-high-risk woman, supplementary screening is advisable because there is a greater likelihood that cancer indeed is there to be detected. For the near-average-risk woman (dense breasts as the only risk factor), neither supplementary screening examination may be justified owing to the very low likelihood that a mammographically occult cancer is present and the increase in false-positive findings that will likely accompany supplementary screening with US or MR imaging. Most difficult to decide is the role of supplementary screening for women at intermediate risk (dense breasts plus one or two major risk factors [eg, a strong family history of breast or ovarian cancer, personal history of breast cancer, or a previous biopsy showing a high-risk lesion]). However, for any given woman at whatever risk, if her decision is to proceed with supplementary screening (the primary goal being earlier cancer detection), MR imaging clearly is the more sensitive test so it should be considered first, with US reserved for extenuating circumstances (MR imaging not available, not tolerated, not affordable, etc) (3,4).

Breast density notification legislation focuses one to consider supplementary screening at the clinical (individual patient) level, but the addition of supplementary screening raises other concerns that should be considered in today's health care environment. In addition to

Published online

10.1148/radiol.13131782 Content code: HP

Radiology 2013; 269:646–647

¹From the Department of Radiology, Emory University Hospital, 1701 Uppergate Dr, Suite C1104, Atlanta, GA 30322. Received and final version accepted July 26, 2013.

Address correspondence to the author (e-mail: carl.dorsi@emoryhealthcare.org).

Conflicts of interest are listed at the end of this article.

See also the article by Price et al in this issue.

© RSNA, 2013

sensitivity and specificity, one also may take into account availability, cost, workforce issues, and likelihood of health insurance coverage for MR imaging versus US—each of which adds substantial complexity to the analysis.

Another equally important issue relates directly to the current controversy surrounding supplementary screening. What is the most meaningful way to compare the observed outcomes of the various available breast imaging modalities? To date, standard auditing procedures have been developed only for mammography as part of BI-RADS. All published data on the other modalities are derived from outcomes analyses that do not rigorously follow standard auditing procedures, potentially confounding observed outcomes. Key to any approach to meaningful auditing is that procedures must be defined precisely, must use only objective (never subjective) criteria, must be applied uniformly (in this case, across all imaging modalities used for breast cancer screening), and must not allow for exceptions. Fortunately, the solution is at hand, spelled out in detail in the new edition of BI-RADS (2), in which a standard auditing approach is described for mammography, MR imaging, and US that uses identical procedures wherever practical. This approach is based on the widely accepted procedures already established for mammography. Specifically as concerns screening (and hence including supplementary screening), a negative test outcome is defined as routine examination of an asymptomatic woman, involving the recording of only a standard set of images, accompanied by a negative (category 1) or benign (category 2) assessment. Complexity is introduced only concerning what represents the standard set of images. For mammography, this is the mediolateral oblique and craniocaudal view of each breast. For MR imaging, this is a facility-designated set of pulse sequences obtained before and after contrast material administration. Although variation in timing and section thickness may be different from facility to facility, the combination of pre- and postcontrast sets of images for breast evaluation is standard. For US,

this may be more complicated owing to the different approaches to examination (scanning only by interpreting physician, only by trained sonographer, or initially by sonographer and then by a physician as needed). To encompass all of these scanning approaches, using only objective criteria applied uniformly without exceptions, BI-RADS recommends as a standard set of images those used in the American College of Radiology Imaging Network 6666 study (4) for negative examinations (one image per quadrant plus one of the retroareolar area for each breast). At screening mammography interpreted while the patient remains at the facility, obtaining only one additional image (other than to overcome technical inadequacy) causes a screening examination to be audited as incomplete (test-positive) and the additional image(s) to be audited as a concurrent diagnostic examination, even though only one examination is performed and only one report is issued. Thus, at screening US, obtaining a full set of images of any specific finding (images in two planes, with and without calipers) has the same auditing result: The screening examination is audited as incomplete (test-positive) and the additional image(s) is audited as a concurrent diagnostic examination, even though only one examination is performed and only one report is issued. Note that BI-RADS auditing does provide the flexibility for characteristically benign findings to be fully assessed at real-time imaging before recording only the standard images (one or more of which could include a characteristically benign finding), recognizing that, just as at screening mammography, there is no need to obtain a full set of diagnostic images for a finding assessed as characteristically benign (category 2). Screening MR imaging is different from both mammography and US in that the standard set of images recorded for a screening examination is the same as is recorded for a diagnostic examination, so that screening MR imaging is essentially also diagnostic MR imaging. This affects MR imaging auditing in that the metrics recall rate and positive predictive value within 1 year from a positive screening examination are essentially

meaningless. Note that BI-RADS does not yet describe auditing procedures for screening with either digital breast tomosynthesis or automated whole-breast US, each of which may (depending on how used) also be equivalent to a diagnostic examination.

The importance of standardized auditing for all imaging modalities is that cross-modality comparisons are accurate and clinically meaningful, which is especially relevant when judging the relative strengths and limitations of supplementary screening with use of the various modalities. It is expected that, with publication of the new edition of BI-RADS, all future clinical research should use auditing procedures (conduct outcomes analyses) that follow the newly established BI-RADS standard and that peer reviewers and journal editors will require use of these procedures as a condition of acceptance for publication.

Acknowledgment: I thank Edward Sickles, MD, for his review of and suggestions for this manuscript.

Disclosures of Conflicts of Interest: Financial activities related to the present article: none to disclose. Financial activities not related to the present article: is a consultant for Philips; owns stock/stock options in Hologic. Other relationships: none to disclose.

References

1. Price E, Hargreaves J, Lipson J, et al. The California Breast Density Information Group: a collaborative response to the issues of breast density, breast cancer risk, and breast density notification legislation. *Radiology* 2013;269(3):887-892.
2. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA. Breast imaging reporting and data system: ACR BI-RADS—breast imaging atlas. 5th ed. Reston, Va: American College of Radiology, 2013 (in press).
3. Lehman CD, Isaacs C, Schnall MD, et al. Cancer yield of mammography, MR, and US in high-risk women: prospective multi-institution breast cancer screening study. *Radiology* 2007;244(2):381-388.
4. Berg WA, Zhang Z, Lehrer D, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA* 2012;307(13):1394-1404.

Breast Density Legislation and Opportunities for Patient-centered Outcomes Research¹

Christoph I. Lee, MD, MSHS
Lawrence W. Bassett, MD
Constance D. Lehman, MD, PhD

One important strategy for reducing breast cancer mortality is early detection through screening (1). Despite a reported decline in mortality rates because of mammography, its effectiveness remains heavily debated. Exemplified by the breast imaging community's backlash against the U.S. Preventive Services Task Force (USPSTF) for its recommendation not to routinely screen women aged 40–49 years, the interpretation of available evidence remains a highly charged and emotional issue for many stakeholders (2).

Not surprisingly, breast cancer screening continues to be one of the most heavily legislated issues in U.S. preventive medicine. Starting with the Mammography Quality Standards Act (MQSA) of 1992 and its reauthorizations in 1998 and 2004, minimum national standards in regard to the operation of mammography equipment, film processing, image interpretation, and results reporting have been instituted. These laws have assured that minimal process measures necessary for decreasing variability in screening practices are being maintained.

Breast Density Screening Legislation

Over the past decade, however, breast imaging has moved beyond the mammogram for women at increased risk. Screening breast ultrasonography (US) and breast magnetic resonance (MR) imaging are two of the available tools that can increase sensitivity for detecting early cancers, especially among women with dense breasts, who may have cancers obscured by large amounts of overlapping fibroglandular tissue. Given the rapid diffusion of these technologies and a movement toward increased shared decision making (as recommended by the USPSTF), there is now a push by patient advocacy groups

for new legislation that would mandate disclosure of breast density information directly to women.

As of April 2012, Connecticut, Texas, and Virginia have adopted such a reporting requirement for women with dense breasts, and at least 10 additional states will consider similar bills in 2012. At a national level, the Breast Density and Mammography Reporting Act (H.R. 1302) was introduced in the 112th U.S. Congress and would require that every mammography report “contain information regarding the patient's breast density and language communicating that individuals with more dense breasts may benefit from supplemental screening tests” (3). If strictly enforced, these recently passed and proposed state and federal laws may drastically change screening practices for women with dense breasts.

Reporting Breast Density to Patients

Women with breast density in the upper quartile have an associated three to five times greater risk of developing breast cancer relative to women with breast density in the lower quartile, even after adjusting for associated risk factors such as age and body mass index (4–7). A previous “masking bias” hypothesis—that the observed higher relative risk was solely due to mass obscuration by dense tissues at mammography—has been debunked by recent large cohort studies (5,8–10). Indeed, breast density is now an established independent risk factor for developing breast cancer, irrespective of the influence of other known risk factors, method of density measurement, or patient population studied (11). Because dense breasts are common, with 31%–43% of the general screening population having heterogeneously dense or extremely dense breasts at mammog-

¹From the Department of Radiology, University of Washington School of Medicine, Seattle, Wash (C.I.L., C.D.L.); Seattle Cancer Care Alliance, 825 Eastlake Ave E, G3-200, Seattle, WA 98109-1023 (C.I.L., C.D.L.); and Department of Radiological Sciences, David Geffen School of Medicine at UCLA, Los Angeles, Calif (L.W.B.). Received January 22, 2012; revision requested March 17; revision received April 10; accepted April 24; final version accepted May 3. Address correspondence to C.I.L. (e-mail: stophlee@uw.edu).

raphy (4), a mandate to disclose the associated increased risk for cancer directly to individual patients seems ethical, reasonable, and appropriate.

Supplemental Screening

The question remains, however, as to what supplemental screening studies patients should be referred to after being informed about their dense breasts. US, given its wide availability and relatively low direct medical costs, is likely the most promising adjunct screening modality currently available. In a recent report (12) from the American College of Radiology Imaging Network (ACRIN) 6666 investigators, the addition of a screening US examination for women with dense breast tissue and at least one other known risk factor resulted in an additional 4.3 cancers detected per 1000 women screened. This finding was commensurate with findings in prior multicenter trials (13–15). In addition, ACRIN 6666 investigators found that 3.7 additional cancers were detected per 1000 women in the second and third rounds of screening (incidence rounds). Across all recent studies, the majority of cancers detected by using additional US are node negative, theoretically leading to earlier treatment for lower-stage invasive cancers and possibly leading to improved patient survival (16,17).

The addition of MR imaging to screening regimens would markedly increase detection of early breast cancer, beyond cancers found at screening mammography alone or at combined screening mammography and US (18–22). MR imaging is unaffected by breast density and, like US, incurs no ionizing radiation. In the recent ACRIN 6666 trial, screening MR imaging was performed in a subset of intermediate-risk women after three negative screening mammography and US examinations. The addition of MR imaging in this patient population yielded the detection of 14.7 additional cancers per 1000 women screened (12). Several studies have reported similar yields from supplemental MR imaging after mammography in the high-risk population (18,20,23,24).

However, additional screening is not without risks to patients. Berg et al (12) found that adjunct US resulted in biopsy in 5% of women in addition to the 2% sent for biopsy on the basis of mammographic findings alone. At these additional biopsies, only 7.4% of the women were found to have cancer. Moreover, while screening US was performed by expert, trained physicians in recent studies, such a practice cannot easily be replicated in the general community given a current manpower shortage. If performed by other personnel, the recall and biopsy rates may become much higher. The addition of screening MR imaging rather than US to mammography in the general community would likely be inappropriate given the current high false-positive rates (19,25). Indeed, 7% of women in the ACRIN 6666 study underwent biopsy on the basis of MR imaging findings alone (12). MR imaging is also less well tolerated by patients, incurs significantly higher costs, is not widely accessible, and includes the risk of adverse events from injection of intravenous contrast material (20). Currently, the American Cancer Society recommends breast MR imaging only in women at high risk for breast cancer and, at this time, considers the evidence insufficient to recommend screening breast MR imaging in women with dense breast tissue but no other risk factors (26).

Current Shortcomings

The advocacy push to legislate mandatory reporting of breast density and possible adjunct screening for all women with heterogeneously or extremely dense breasts is far outpacing the reporting of evidence that supplemental screening may provide better outcomes for these patients. Recent study results in regard to adjunct screening US and MR imaging, while encouraging, pertain only to women of intermediate or high risk, with known risk factors beyond their dense breasts. Therefore, it is uncertain what the added cancer detection yield of supplemental screening would be for women of average risk with dense breasts but no other known risk factors.

Even with increased rates of early cancer detection, the impact of supplemental screening on patient morbidity and mortality remains unknown. Recent trials did not include control groups, meaning that the impact of additional screening on patient mortality cannot be determined (12). Beyond survival benefit, the question arises as to whether detection of more abnormalities will lead to increases in overdiagnosis. Because some cancers detected at screening may not go on to cause symptoms or death, additional interventions performed on these excess cancers would only increase morbidity for these patients (27).

If the demand for supplemental screening increases at a high rate, then issues with supply will have to be addressed. Currently, there is a shortage of qualified breast imagers and breast US technologists who can perform competent screening US examinations (12). Rates of cancer detection at technologist-performed screening US appear to be similar to those at physician-performed US, but a large investment would have to be made to train more technologists (28). Automated whole-breast US promises to decrease operator variability, but this technology has just received Food and Drug Administration approval, and its accuracy for depicting smaller cancers is yet to be determined (29,30).

It is also uncertain who will pay for additional screening if recommended by law. Currently, only two states—Illinois and Connecticut—have considered mandating insurance coverage for supplemental screening US. In Illinois, “if a routine mammogram reveals heterogeneous or dense breast tissue, coverage must provide for a comprehensive ultrasound screening of an entire breast or breasts, when determined to be medically necessary by a physician” (31). Even if covered by insurance, there is cur-

Published online
10.1148/radiol.12120184

Radiology 2012; 264:632–636

Potential conflicts of interest are listed at the end of this article.

rently only one Current Procedural Terminology code available for breast US, with a low Medicare reimbursement level (approximately \$90) that may not adequately cover the cost of the physician time required for performing and interpreting a comprehensive study (12). To date, from a health systems standpoint, there are no randomized control trials or cost-effectiveness analyses demonstrating that supplemental screening is a cost-effective measure for women with dense breasts.

Within the highly litigious environment of breast cancer screening, it is not unreasonable to expect an increase in reflexive ordering of unnecessary supplemental studies for women with dense breasts but no other known risk factors. If there is a legislated recommendation that the patient may benefit from additional screening, then an order for supplemental screening devoid of an individual patient-centered risk-benefit discussion may result because of the physician's concern for medical-legal protection. Such reflexive ordering would lead to increased inappropriate utilization of breast imaging technologies at increased costs and decreased net health benefits.

Building a Framework for Patient-centered Outcomes Research

Given these current shortcomings, what is needed is a unified, organized approach to building a framework for identifying and addressing the key issues in regard to determining the effects of adjunct screening on individual outcomes for women with dense breasts. The maelstrom encompassing the proposed breast density legislation, the recent USPSTF recommendations, and the Patient Protection and Affordable Care Act (PPACA) has created an unusual and fortuitous climate for change in national breast screening practices. We argue that building such a framework centered on comparative effectiveness research (CER) and patient-centered outcomes research is critical not only for addressing the needs created by new density legislation but also for collecting the evidence that will ultimately

best inform individual risk-based discussions between patients and health care providers regarding breast cancer screening. In the remaining paragraphs, we introduce three core issues that will need to be addressed up front—consensus, quality, and cost.

Consensus

CER and patient-centered outcomes research promise to inform health care decisions by providing evidence on the effectiveness, benefits, and harms of different screening options for different patients. However, before key research initiatives can be identified, all key stakeholders must come together to build consensus for a shared research agenda. Health care reform, culminating in the PPACA, has established the Patient-Centered Outcomes Research Institute (PCORI), an independent organization created to help people make informed health care decisions and improve health care delivery (32). According to the PCORI, patient-centered outcomes research is “guided by patients, caregivers and the broader health care community and will produce high integrity, evidence-based information.” Thus, one of the main tenets of research in the era of health care reform will be the heavy involvement of patients at each stage of research. In fact, much of the current drive for breast density legislation likely stems from the historic lack of communication between the medical community and its patients in regard to the limitations of mammography. The breast imaging community, therefore, must welcome patient advocacy groups into the fold, along with other key stakeholders, including payer organizations.

Quality

Before performing comparative effectiveness studies involving adjunct technologies for breast cancer screening, all stakeholders collectively must establish mandatory minimum quality standards for newer modalities, similar to the MQSA for mammography. The subjectivity and variation in breast imaging remain beyond mammography, and satisfactory process measures must be created for the operation, maintenance,

image processing, and reporting of screening breast US. Contrary to popular belief within the medical community, mandatory accreditation and certification of imaging facilities for breast US are not currently required by federal law. While the American College of Radiology (ACR) has breast US and US-guided biopsy accreditation and certification programs (33), these are for the most part optional. Establishing mandatory accreditation may provide a level of standardization necessary for ensuring a minimum level of competency in process measures and allow for the evaluation of the comparative effectiveness of different modalities and screening strategies. Moreover, inter- and intraobserver variability in interpretation for new screening modalities must be addressed, with the development of methods to improve standardization in physician interpretation (34). Practitioners should meet certain experience and continuing education criteria for performance and interpretation of screening US studies set by the ACR or the American Institute for Ultrasound in Medicine (16).

Cost

The cost associated with any new screening strategy is critical and should be dealt with up front, in parallel with efforts to determine improved patient outcomes. Health care reform demands that new interventions be of higher value for lower costs to patients, the health care system, and society. Randomized controlled trials in which different screening strategies are compared for women with dense breasts are unlikely to be performed because of the associated large number of patients needed to demonstrate a difference between groups, the long length of follow-up required, and the large monetary expenses incurred. In the absence of definitive randomized controlled trials to establish the comparative effectiveness of multimodality breast cancer screening, computer simulation models of breast cancer natural history and outcomes can be used to project long-term health outcomes and lifetime costs related to different screening strategies.

Therefore, while long-term prospective studies through ACRIN and other research collaboratives should be pursued, preliminary decision analyses and cost-effectiveness analyses must be performed with models based on available efficacy data and expert opinion to guide current decision making. Historically, in the United States, the key driver of widespread use of new breast imaging interventions has been reimbursement by government and third-party payers rather than the reporting of clinical efficacy (35). Thus, payers must be engaged as partners early to help establish appropriate reimbursement rates and increase access to new adjunct screening tools. Government payers, for instance, may be able to provide coverage for supplemental screening in return for requiring evidence collection, as is the case for positron emission tomography (PET) and Medicare through the national PET registry.

With engaged patient advocacy groups, research should include aspects of overall costs important to individual patients, such as out-of-pocket costs, transient levels of anxiety from false-positive findings or biopsies, and lost time for follow-up. Patients can help clarify their preferences and values, and the resulting quality-of-life utilities should be included in all analyses. Furthermore, cost-effectiveness analyses should be focused on specific subpopulations of patients, such as women with dense breasts of average risk and no other risk factors, and should compare specific strategies, such as combined mammography and US screening versus mammography alone. In an era of more personalized breast cancer care, the cost-effectiveness of a specific screening strategy may be dependent on an individual's specific risks of developing breast cancer (36–38). Our analyses should reflect this trend toward individualized breast cancer care and should truly inform the personal decision-making process.

Conclusion

The timely convergence of advocacy efforts, high political will, and health care

reform provides an important opportunity for the breast cancer community to help institute positive change in screening practices. However, current legislation that mandates informing patients about possible increased breast cancer risk on the basis of breast density may not suffice in actually improving patient outcomes. Instead, the breast imaging community must partner with advocacy groups to shift efforts toward creating an infrastructure for patient-centered outcomes research that will provide the evidence needed for meaningful discussions between individuals and their physicians about screening. Through early consensus building, standardized quality measures, and a focus on cost effectiveness, we can maximize the benefits of breast cancer screening for every woman. Moreover, by partnering with all stakeholders at each stage of our research efforts, we can create an increased level of transparency to help prevent future controversies in breast cancer screening recommendations.

Disclosures of Potential Conflicts of Interest:

C.I.L. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: is a consultant for GE in their breast imaging workshop; institution receives money from a GE–Association of University Radiologists academic research award. Other relationships: none to disclose. **L.W.B.** No potential conflicts of interest to disclose. **C.D.L.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: is a consultant for Philips, GE, and Bayer; institution has grants or grants pending from GE; receives and institution receives payment for development of educational presentations from GE. Other relationships: none to disclose.

References

1. Nyström L, Andersson I, Bjurström N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 2002;359(9310):909–919.
2. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2009;151(10):716–726, W-236. [Published correction appears in *Ann Intern Med* 2010;152(10):688 and *Ann Intern Med* 2010;152(3):199–200].
3. H.R. 1302. Breast Density and Mammography Reporting Act of 2011. U.S. Congress. <http://www.govtrack.us/congress/bills/112/1302>. Published October 2011. Accessed November 15, 2011.
4. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006;15(6):1159–1169.
5. Byrne C, Schairer C, Wolfe J, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *J Natl Cancer Inst* 1995;87(21):1622–1629.
6. Boyd NF, Rommens JM, Vogt K, et al. Mammographic breast density as an intermediate phenotype for breast cancer. *Lancet Oncol* 2005;6(10):798–808.
7. Kerlikowske K, Ichikawa L, Miglioretti DL, et al. Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk. *J Natl Cancer Inst* 2007;99(5):386–395.
8. Maskarinec G, Pagano I, Lurie G, Kolonel LN. A longitudinal investigation of mammographic density: the multiethnic cohort. *Cancer Epidemiol Biomarkers Prev* 2006;15(4):732–739.
9. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356(3):227–236.
10. Boyd NF, Byng JW, Jong RA, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *J Natl Cancer Inst* 1995;87(9):670–675.
11. Vachon CM, van Gils CH, Sellers TA, et al. Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Res* 2007;9(6):217.
12. Berg WA, Zhang Z, Lehrer D, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA* 2012;307(13):1394–1404.
13. Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA* 2008;299(18):2151–2163.
14. Corsetti V, Ferrari A, Ghirardi M, et al. Role of ultrasonography in detecting mammographically occult breast carcinoma in women with dense breasts. *Radiol Med (Torino)* 2006;111(3):440–448.
15. Tohno E, Ueno E, Watanabe H. Ultrasound screening of breast cancer. *Breast Cancer* 2009;16(1):18–22.
16. Berg WA. Supplemental screening sonography in dense breasts. *Radiol Clin North Am* 2004;42(5):845–851, vi.

17. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AM, Chen TH. The randomized trials of breast cancer screening: what have we learned? *Radiol Clin North Am* 2004;42(5):793–806, v.
18. Kuhl CK, Schrading S, Leutner CC, et al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *J Clin Oncol* 2005;23(33):8469–8476.
19. Kuhl C, Weigel S, Schrading S, et al. Prospective multicenter cohort study to refine management recommendations for women at elevated familial risk of breast cancer: the EVA trial. *J Clin Oncol* 2010;28(9):1450–1457.
20. Warner E, Plewes DB, Hill KA, et al. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA* 2004;292(11):1317–1325.
21. Sardanelli F, Podo F, Santoro F, et al. Multicenter surveillance of women at high genetic breast cancer risk using mammography, ultrasonography, and contrast-enhanced magnetic resonance imaging (the high breast cancer risk Italian 1 study): final results. *Invest Radiol* 2011;46(2):94–105.
22. Vacek PM, Geller BM. A prospective study of breast cancer risk using routine mammographic breast density measurements. *Cancer Epidemiol Biomarkers Prev* 2004;13(5):715–722.
23. Lehman CD, Isaacs C, Schnall MD, et al. Cancer yield of mammography, MR, and US in high-risk women: prospective multi-institution breast cancer screening study. *Radiology* 2007;244(2):381–388.
24. Sardanelli F, Podo F, D'Agnolo G, et al. Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRIT study): interim results. *Radiology* 2007;242(3):698–715.
25. Lehman CD, Smith RA. The role of MRI in breast cancer screening. *J Natl Compr Canc Netw* 2009;7(10):1109–1115.
26. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007;57(2):75–89.
27. Elmore JG, Fletcher SW. Overdiagnosis in breast cancer screening: time to tackle an underappreciated harm. *Ann Intern Med* 2012;156(7):536–537.
28. Kaplan SS. Clinical utility of bilateral whole-breast US in the evaluation of women with dense breast tissue. *Radiology* 2001;221(3):641–649.
29. Kelly KM, Dean J, Comulada WS, Lee SJ. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *Eur Radiol* 2010;20(3):734–742.
30. Lander MR, Tabár L. Automated 3-D breast ultrasound as a promising adjunctive screening tool for examining dense breast tissue. *Semin Roentgenol* 2011;46(4):302–308.
31. Illinois Department of Insurance. Coverage for the Detection and Treatment of Breast Cancer. http://insurance.illinois.gov/health_insurance/breastcancer.asp. Published July 2009. Accessed March 30, 2012.
32. Patient-Centered Outcomes Research Institute. Overview. <http://www.pcori.org/>. Accessed March 29, 2012.
33. American College of Radiology. Ultrasound Accreditation Program. <http://www.acr.org/accreditation/ultrasound.aspx>. Accessed March 30, 2012.
34. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR Am J Roentgenol* 2000;174(6):1769–1777.
35. Gold LS, Klein G, Carr L, Kessler L, Sullivan SD. The emergence of diagnostic imaging technologies in breast cancer: discovery, regulatory approval, reimbursement, and adoption in clinical guidelines. *Cancer Imaging* 2012;12(1):13–24.
36. Darabi H, Czene K, Zhao W, Liu J, Hall P, Humphreys K. Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. *Breast Cancer Res* 2012;14(1):R25.
37. Katz SJ, Morrow M. The challenge of individualizing treatments for patients with breast cancer. *JAMA* 2012;307(13):1379–1380.
38. Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Ann Intern Med* 2011;155(1):10–20.

To Seek Perfection or Not? That Is the Question

Carl J. D'Orsi, MD
Edward A. Sickles, MD

The article in this month's issue of *Radiology* (1), concerning ultrasonography (US) and cancer detection in women with dense breasts and the recently enacted legal mandates in Connecticut, are topics that we, as breast imagers, must understand. Connecticut Public Act 09-41, section 2, section 38a-530 (c), states:

On and after October 1, 2009, each mammography report provided to a patient *shall* [emphasis added] include information about breast density, based on the Breast Imaging Reporting and Data System established by the American College of Radiology. Where applicable, such report *shall* [emphasis added] include the following notice: If your mammogram demonstrates that you have dense breast tissue, which could hide small abnormalities, you might benefit from supplementary screening tests, which can include a breast ultrasound screening or a breast MRI examination, or both, depending on your individual risk factors. A report of your mammography results, which contains information about your breast density, has been sent to your physician's office and you should contact your physician if you have any questions or concerns about this report.

We have italicized "shall" since in legislative language this is equivalent to "must" in everyday language.

The report by Hooley and colleagues presents outcomes data on what they choose to call "screening" breast US in women with dense breasts, performed in Connecticut after passage of Public Act 09-41. It is clear that the data presented do not represent the established definition of "screening." In the article by Hooley et al (1), the population of women who underwent a screening US examination were both from a true screening population (asymptomatic with negative findings at

mammography), as well as a diagnostic population (symptomatic and/or with positive findings at mammography). The universally accepted definitions of *screening* and *diagnostic* for breast imaging examinations are critical to maintain. This allows us to perform valid comparisons of breast imaging technologies and critically assess whether their greatest impact would be in a screening or diagnostic setting or both.

The breast cancer detection rate and cost per cancer detected reported by Hooley et al (1) are also affected by their mixture of patients who underwent screening and those who underwent diagnostic examinations. Given their mixed population (which has a higher prior probability of malignancy), one would expect more cancer detection than for a true screening population, producing a falsely elevated cancer detection rate for screening breast US. This mixture would also falsely reduce the cost per cancer detected. The authors state that women with dense breast tissue, as reported from the mammogram, received the statement about supplementary screening required by Connecticut law. However, it is also stated that a direct determination of the number of studied women with dense breasts could not be done. It thus remains unclear exactly how these women were identified. Currently the definition of "dense" breasts is a subjective rating by the interpreter, as defined in Breast Imaging Reporting and Data System (BI-RADS), which in Connecticut includes the categories of "heterogeneously dense" and "extremely dense." A much more valid assessment would be the volume of dense tissue as a percentage of total breast volume. This cannot be accurately estimated by orthogonal two-dimensional images (2) and probably would require a standardized true isotropic three-dimensional (3D) capture of the whole breast. We are not there yet.

Published online before print

10.1148/radiol.12121515 Content code: **BR**

Radiology 2012; 265:9-11

¹From the Department of Breast Imaging Research, Emory University Hospital, WCI Building, 1365-C Clifton Rd, Suite C1104, Atlanta, GA 30322 (C.J.D.); and Department of Radiology, University of California San Francisco School of Medicine, San Francisco, Calif (E.A.S.). Received July 10, 2012; final version accepted July 17. **Address correspondence to C.J.D.** (e-mail: cdorsi@emory.edu).

Potential conflicts of interest are listed at the end of this article.

See also the article by Hooley et al in this issue.

© RSNA, 2012

The mandated language in Connecticut Public Act 09-41 raises another problem. It mentions the difficulty for cancer detection in dense breasts and the increased risk for developing breast cancer in women with dense breasts. These are two separate matters. Intuitively, it is not difficult to understand limited detection and reduced sensitivity in dense breasts. There are reports describing this phenomenon. One such report (3) notes sensitivity for cancer detection of 30% in women with dense breasts versus 80% in women without dense breasts. They also describe an increase in the interval cancer rate, which is a by-product of the lower sensitivity. Breast density as a risk factor for development of breast cancer is far more problematic. The subjective evaluation of breast density certainly has a role in the calculation of the risk for breast cancer. However, there are many confounding variables that are difficult to isolate, even with sophisticated statistical analysis. Furthermore, many publications report a misleadingly high relative risk for breast density by using as a referent group the small percentage of women with almost entirely fatty breasts instead of women of average breast density (4). While an exhaustive discussion of this topic is not the point of this editorial, a simple example may suffice. Does a woman classified as not having dense breasts visually, and thus is at lower risk for breast cancer, increase her risk if she loses weight and fatty tissue from her breasts, which then are evaluated as dense? This is truly problematic since obesity has been associated with increased breast cancer risk (5).

Metrics for both screening US and magnetic resonance (MR) imaging are relatively newer because these breast imaging modalities have only recently been applied to screening. The definitions for the four basic terms (true-positive [TP] findings, false-positive [FP] findings, true-negative [TN] findings, false-negative [FN] findings) that are used for calculation of sensitivity, specificity, and accuracy are most mature and accepted for screening mammography. We must strictly adhere to

the definitions of these parameters for the other breast imaging modalities, on the basis of similar principles, if we are to correctly assess their effect. The underlying assumption for screening mammography is the separation of a screening from a diagnostic examination with different sensitivity, specificity, and accuracy results. The screening examination consists of a limited set of images (for a bilateral examination, four views, occasionally with additional views to cover all the breast tissue or correct technologist-detected artifacts). It is frequently interpreted in a batch mode at a time separate from acquisition. An examination with negative or benign findings is defined as an examination with the presence of nothing of concern or one with definitely benign findings, with a recommendation for a routine screening in a year. The diagnostic examination involves women with a clinical or screening-detected area of concern, and supplemental mammographic views are added and frequently US is added.

The new edition of ACR BI-RADS (fifth edition) (6) emphasizes that a probably benign assessment (category 3, with recommendation for short-interval follow-up) not be used at screening. Interpreters will now pay a penalty in potential FPs because this new edition defines the auditing of assessments with findings classified as category 3 at screening as examinations with positive findings, similar to assessments with findings classified as category 0. This change reflects the rationale that findings classified as category 3 at screening basically indicate a recommendation for "work-up" in the form of additional imaging prior to the next routine screening examination, albeit 6 months rather than promptly after the index examination. We should apply the same definition for screening breast US as we do for screening mammography: an examination with a limited number of defined views, read as having positive (further action needed prior to the next routine screening examination) or negative (nothing needed until the next routine screening examination) findings. However, screening US is different

because it has real-time imaging capability. This difference requires us to address this facet of breast US in the definitions that go into sensitivity, specificity, and accuracy.

Screening breast US may be accomplished by a technologist, a radiologist, or an automated system. When obtained by a technologist or an automated system, routine "snapshots" may be recorded with perhaps an included standard, routine cine loop to utilize real-time capabilities. Then after interpretation by using these two methods of acquisition, any request for additional examination, including the radiologist performing directed scanning of one or more target areas, should be considered a screening examination with positive findings. The more complicated scenario involves screening performed initially by the radiologist, who then may utilize additional US maneuvers during the screening to resolve any problems raised. Thus, the interpreter is acting both as screener and diagnostic evaluator. One way to address this issue in an effort to standardize breast US screening metrics is to define as routine the "snapshot" images of the breast quadrants and retroareolar area obtained by the radiologist, and if any further images are recorded, the US examination would then be considered a screening examination with positive findings, followed by a diagnostic examination. This is the same approach that is applied to mammography, but instead of separating screening and diagnostic examinations by days, they are separated by minutes.

This approach does not apply to MR imaging because whether the examination is performed as a screening examination or for diagnostic purposes, an identical set of images is recorded, so that MR imaging is truly acting simultaneously as both a screening and a diagnostic examination.

The above-described approaches to assessing results of screening mammography, US, and MR imaging have been incorporated into the fifth edition of BI-RADS. This is vital if we are to meaningfully assess newer versus established

modalities by using metrics that are universally understood and similarly applied.

Accepting the newer BI-RADS approaches in regard to metric definitions, when considering the article on screening breast US by Hooley et al (1), we can relook at the presented data. The cohort who underwent screening breast US was comprised of 935 women: In 701 women, findings were assessed as BI-RADS category 1 or 2; in 187 women, findings were assessed as BI-RADS category 3; and in 47 women, findings were assessed as BI-RADS category 4. Three cancers were detected. How did screening US with positively interpreted results perform as a predictor of breast cancer? Considering all the results classified as BI-RADS category 3 and 4 assessments at screening as positive, the percentage of all positive findings of screening examinations that result in a tissue diagnosis of cancer within a year (positive predictive value [PPV]₁) is 1.3% (three of 234). Using the same approach, the PPV₁ of screening mammography ranges from 4% to 9% (7). Both sets of metrics would rely on the demographics of the population tested (ie, age, race, socioeconomic status). When used as a diagnostic examination, US results were indicative of a recommendation for biopsy and US was used to achieve biopsies in 46 women. Thus, PPV for biopsy performed (PPV₃) is 6.5% (three of 46). The PPV₃ for diagnostic mammography is 39.5% (8). What is difficult to determine and not described by Hooley et al (1) is the performance of screening breast US as an adjunct to mammography performed at the same time. The American College of Radiology Imaging Network, or ACRIN, 6666 trial gives us some insight, reporting that mammography alone had a PPV₁ of 9.6% and in concert with US a PPV₁ of 7% (9). However, for detection of breast cancer, the sensitivity for mammography

alone was 31.3% but combined with US was 43.8%. How do we analyze these data? FN and TP findings drive sensitivity, but FN findings tend to have a greater effect. What drives PPV₁ is TP and FP findings, but FP findings for this metric are more critical. The receiver operating characteristic informs us of two important relationships: As FP findings increase, FN findings decrease. However, above a certain threshold level of cancer detection we pay a very high FP price to attain fewer FN findings and more cancer detection.

The lack of separation between screening and diagnostic examinations in the report of Hooley et al (1) has considerable relevance as a potential indicator of future benchmark performance, given that the reported data skew the outcomes expected from true screening US to falsely appear more favorable than they really are. These data may be used to affect the decisions of lawmakers who are considering similar legislation in other states. In our opinion, such government mandates are premature. Undoubtedly, it is beneficial to more fully inform women about their breast health; the pertinent question is how to do this. Even if one is willing to assume that early cancer detection by using screening US will translate into breast cancer mortality reduction, does this benefit of supplementary screening exceed the harms, given that compared with screening mammography, the TP rate of supplementary screening US is lower and the FP rate is much higher (TP rate even lower and FP rate even higher than those reported by Hooley et al [1])? If a governmental mandate is established to inform women with dense breasts of the benefits and harms of supplementary screening, shouldn't insurance coverage for this additional testing also be mandated? Specifically, why inform without providing the means to follow through, knowing that absent a mandate for insurance coverage, only the more affluent will be able

to afford the cost of supplementary screening US?

Disclosures of Potential Conflicts of Interest:

C.J.D. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: paid consultant to Fuji and Phillips; institution received grants from Fuji; received stock or stock options from Hologic. Other relationships: none to disclose. **E.A.S.** No potential conflicts of interest to disclose.

References

1. Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. *Radiology* 2012;265(1):59-69.
2. Kopans DB. Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. *Radiology* 2008;246(2):348-353.
3. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* 2000;92(13):1081-1087.
4. Sickles EA. The use of breast imaging to screen women at high risk for cancer. *Radiol Clin North Am* 2010;48(5):859-878.
5. Kerlikowske K, Walker R, Miglioretti DL, Desai A, Ballard-Barbash R, Buist DS. Obesity, mammography use and accuracy, and advanced breast cancer risk. *J Natl Cancer Inst* 2008;100(23):1724-1733.
6. D'Orsi CJ, Mendelson EB, Morris EA, et al. ACR BI-RADS: Breast Imaging Reporting and Data System. 5th ed. Reston, Va: American College of Radiology, 2012 (in press).
7. Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. *Radiology* 2006;241(1):55-66.
8. Sickles EA, Miglioretti DL, Ballard-Barbash R, et al. Performance benchmarks for diagnostic mammography. *Radiology* 2005;235(3):775-790.
9. Berg WA, Zhang Z, Lehrer D, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA* 2012;307(13):1394-1404.



Scan for Author
Audio Interview

Detection of Breast Cancer With Addition of Annual Screening Ultrasound or a Single Screening MRI to Mammography in Women With Elevated Breast Cancer Risk

Wendie A. Berg, MD, PhD

Zheng Zhang, PhD

Daniel Lehrer, MD

Roberta A. Jong, MD

Etta D. Pisano, MD

Richard G. Barr, MD, PhD

Marcela Böhm-Vélez, MD

Mary C. Mahoney, MD

W. Phil Evans III, MD

Linda H. Larsen, MD

Marilyn J. Morton, DO

Ellen B. Mendelson, MD

Dione M. Farria, MD

Jean B. Cormack, PhD

Helga S. Marques, MS

Amanda Adams, MPH

Nolin M. Yeh, MS

Glenna Gabrielli, BS

for the ACRIN 6666 Investigators

SIX PREVIOUS SINGLE-CENTER studies¹⁻⁶ and 3 multicenter trials⁷⁻⁹ have shown supplemental screening breast ultrasound significantly increases detection of node-negative invasive breast cancer in women with mammographically dense breast tissue on the first prevalence screen, consistently increasing cancer detection (yield) by 3.5 per 1000 in single-center studies and from 4.2 to 4.4

See also p 1379.

Author Video Interview available at www.jama.com.

Context Annual ultrasound screening may detect small, node-negative breast cancers that are not seen on mammography. Magnetic resonance imaging (MRI) may reveal additional breast cancers missed by both mammography and ultrasound screening.

Objective To determine supplemental cancer detection yield of ultrasound and MRI in women at elevated risk for breast cancer.

Design, Setting, and Participants From April 2004-February 2006, 2809 women at 21 sites with elevated cancer risk and dense breasts consented to 3 annual independent screens with mammography and ultrasound in randomized order. After 3 rounds of both screenings, 612 of 703 women who chose to undergo an MRI had complete data. The reference standard was defined as a combination of pathology (biopsy results that showed in situ or infiltrating ductal carcinoma or infiltrating lobular carcinoma in the breast or axillary lymph nodes) and 12-month follow-up.

Main Outcome Measures Cancer detection rate (yield), sensitivity, specificity, positive predictive value (PPV3) of biopsies performed and interval cancer rate.

Results A total of 2662 women underwent 7473 mammogram and ultrasound screenings, 110 of whom had 111 breast cancer events: 33 detected by mammography only, 32 by ultrasound only, 26 by both, and 9 by MRI after mammography plus ultrasound; 11 were not detected by any imaging screen. Among 4814 incidence screens in the second and third years combined, 75 women were diagnosed with cancer. Supplemental incidence-screening ultrasound identified 3.7 cancers per 1000 screens (95% CI, 2.1-5.8; $P < .001$). Sensitivity for mammography plus ultrasound was 0.76 (95% CI, 0.65-0.85); specificity, 0.84 (95% CI, 0.83-0.85); and PPV3, 0.16 (95% CI, 0.12-0.21). For mammography alone, sensitivity was 0.52 (95% CI, 0.40-0.64); specificity, 0.91 (95% CI, 0.90-0.92); and PPV3, 0.38 (95% CI, 0.28-0.49; $P < .001$ all comparisons). Of the MRI participants, 16 women (2.6%) had breast cancer diagnosed. The supplemental yield of MRI was 14.7 per 1000 (95% CI, 3.5-25.9; $P = .004$). Sensitivity for MRI and mammography plus ultrasound was 1.00 (95% CI, 0.79-1.00); specificity, 0.65 (95% CI, 0.61-0.69); and PPV3, 0.19 (95% CI, 0.11-0.29). For mammography and ultrasound, sensitivity was 0.44 (95% CI, 0.20-0.70, $P = .004$); specificity 0.84 (95% CI, 0.81-0.87; $P < .001$); and PPV3, 0.18 (95% CI, 0.08 to 0.34; $P = .98$). The number of screens needed to detect 1 cancer was 127 (95% CI, 99-167) for mammography; 234 (95% CI, 173-345) for supplemental ultrasound; and 68 (95% CI, 39-286) for MRI after negative mammography and ultrasound results.

Conclusion The addition of screening ultrasound or MRI to mammography in women at increased risk of breast cancer resulted in not only a higher cancer detection yield but also an increase in false-positive findings.

Trial Registration clinicaltrials.gov Identifier: NCT00072501

JAMA. 2012;307(13):1394-1404

www.jama.com

per 1000 in multicenter trials. The vast majority of cancers detected only by ultrasound have been node-negative invasive breast cancers. Until now, it was

Author Affiliations and a complete list of the ACRIN 6666 Investigators appear at the end of this article.

Corresponding Author: Wendie A. Berg, MD, PhD, Department of Radiology, Magee-Womens Hospital, 300 Halket St, Pittsburgh, PA 15213 (wendieberg@gmail.com).

unclear whether continuing ultrasound screening annually (ie, incidence screening) would result in a detection benefit.

A substantial majority of American College of Radiology Imaging Network (ACRIN) 6666 participants were at intermediate risk for breast cancer, with more than half having a personal history of breast cancer.⁷ Although there was evidence from prior studies¹⁰⁻¹³ that magnetic resonance imaging (MRI) provided considerable detection benefit beyond what combined ultrasound and mammography screens could discover in high-risk women, the combination of ultrasound and mammography might still identify the vast majority of cancers when they are node negative at a much lower cost to the health care system than the cost of an MRI, particularly when screening women with a lower prevalence of disease. A substudy of ACRIN 6666 participants was therefore undertaken to assess the rate and stage of cancers detected with a single screening MRI.

METHODS

Study Design

Study participants included women who were asymptomatic, presenting for routine annual mammography with heterogeneously dense or extremely dense breast tissue,¹⁴ and who had at least 1 other risk factor for breast cancer (TABLE 1). Race/ethnicity was self-assigned based on fixed categories.

Each participant underwent mammographic and physician-performed ultrasonographic screening examinations in randomized order, with the interpreting radiologist for each examination masked to results of the other study, at 0 months (first screening), 12 months (second screening), and 24 months (third screening). The randomization process has been previously described,⁷ and initial randomization order was maintained for subsequent screening rounds. If recommendation from either screening test was other than routine annual screening, the test was considered positive for purposes of

Table 1. Participant Characteristics

| | Screening Analysis Set, No. (%) | | | |
|---|---------------------------------|------------------------------|-----------------|------------------|
| | 1 (n = 2659) | 2 (n = 2493) ^a | 3 (n = 2321) | MRI (n = 612) |
| Age at scan, mean (SD), y | 55.2 (10.1) | 56.4 (9.9) | 57.7 (9.8) | 56.8 (9.5) |
| Median (range) | 55.0 (25-91) | 56.0 (26-92) | 57.0 (27-93) | 57.0 (27-87) |
| Age group at scan, y | | | | |
| <40 | 134 (5.0) | 89 (3.6) | 65 (2.8) | 17 (2.8) |
| 40-49 | 627 (23.6) | 514 (20.6) | 392 (16.9) | 114 (18.6) |
| 50-69 | 1678 (63.1) | 1644 (65.9) | 1597 (68.8) | 429 (70.1) |
| >69 | 220 (8.3) | 246 (9.9) | 267 (11.5) | 52 (8.5) |
| Race/ethnicity | | | | |
| White | 2467 (92.8) | 2316 (92.9) | 2162 (93.1) | 576 (94.1) |
| Hispanic or Latino | 265 (10.0) | 233 (9.3) | 209 (9.0) | 83 (13.6) |
| Black or African American | 91 (3.4) | 85 (3.4) | 77 (3.3) | 11 (1.8) |
| Native Hawaiian or other Pacific Islander | 4 (0.2) | 3 (0.1) | 4 (0.2) | 1 (0.2) |
| Asian | 90 (3.4) | 82 (3.3) | 71 (3.1) | 22 (3.6) |
| American Indian or Alaskan Native | 4 (0.2) | 4 (0.2) | 4 (0.2) | 1 (0.2) |
| Unknown | 11 (0.4) | 11 (0.4) | 11 (0.5) | 1 (0.2) |
| Menopausal status | | | | |
| Premenopausal ^b | 609 (22.9) | 554 (22.2) | 502 (21.6) | 155 (25.3) |
| Perimenopausal ^c | 182 (6.8) | 170 (6.8) | 158 (6.8) | 37 (6.0) |
| Postmenopausal ^d | 1362 (51.2) | 1294 (51.9) | 1208 (52.0) | 316 (51.6) |
| Surgical menopause | 484 (18.2) | 454 (18.2) | 432 (18.6) | 103 (16.8) |
| Unknown | 22 (0.8) | 21 (0.8) | 21 (0.9) | 1 (0.2) |
| Personal history of breast cancer (regardless of other risk factors) ^e | 1426 (53.6) | 1331 (53.4) | 1253 (54.0) | 275 (44.9) |
| Visually estimated breast density at scan, % | | | | |
| ≤25 | 47 (1.8) | 47 (1.9) | 34 (1.5) | 7 (1.1) |
| 26-40 | 278 (10.5) | 236 (9.5) | 196 (8.4) | 61 (10.0) |
| 41-60 | 824 (31.0) | 792 (31.8) | 774 (33.3) | 191 (31.2) |
| 61-80 | 994 (37.4) | 976 (39.1) | 920 (39.6) | 253 (41.3) |
| >80 | 515 (19.4) | 442 (17.7) | 395 (17.0) | 100 (16.3) |
| Unknown | 1 (<1) | 0 | 2 (0.1) | 0 |
| Primary risk factor ^f | | | | |
| Mutation in <i>BRCA1</i> or <i>BRCA2</i> | 23 (0.9) | 20 (0.8) | 18 (0.8) | 3 (0.5) |
| History of prior chest, mediastinal, or axillary irradiation | 8 (0.3) | 6 (0.2) | 6 (0.3) | 2 (0.3) |
| Personal history of breast cancer | 1413 (53.1) | 1321 (53.0) | 1244 (53.6) | 273 (44.6) |
| Lifetime risk, Gail/Claus model ≥25% ^g | 504 (19.0) | 460 (18.5) | 425 (18.3) | 135 (22.1) |
| 5-Year risk, Gail model ≥2.5% | 406 (15.3) | 391 (15.7) | 366 (15.8) | 113 (18.5) |
| 5-Year risk, Gail model ≥1.7% and extremely dense breasts | 225 (8.5) | 217 (8.7) | 195 (8.4) | 70 (11.4) |
| ADH/ALH/LCIS or atypical papilloma | 80 (3.0) | 78 (3.1) | 67 (2.9) | 16 (2.6) |

Abbreviations: ADH, atypical ductal hyperplasia; ALH, atypical lobular hyperplasia; LCIS, lobular carcinoma in situ; MRI, magnetic resonance imaging.
^aThree participants excluded from year-1 analysis were included in year 2, for a total of 2662 unique participants.
^bDefined as last menstrual period within prior 30 days.
^cDefined as last menstrual period more than 30 days and less than 12 months prior.
^dDefined as last menstrual period at least 12 months prior.
^eEight hundred sixty-nine of 1426 women (71.1%) with a personal history of breast cancer had lumpectomy and radiation therapy for the affected breast(s) during the study.
^fParticipants with multiple risk factors were determined to have a primary risk factor using the following hierarchy: Mutation in *BRCA1* or *BRCA2* genes; history of prior chest, mediastinal, axillary irradiation or all 3; a personal history of breast cancer; a lifetime risk, Gail model of at least 25%; 5-year risk, Gail model of 2.5% or more; 5-year risk, Gail model 1.7% or more; extremely dense breasts; and prior biopsy showing ADH, ALH, LCIS, or atypical papilloma.
^gOne participant's eligibility is based on a recalculated Gail score, for which the original score was missing.

analysis and a qualified site investigator then recorded an integrated interpretation by reviewing study mammogram and ultrasound together. Clinical management was based on integrated interpretation. If both modalities recommended routine annual follow-up, no integration was performed. *Cancers positive only on a given modality* refers to those not visible on any other modality. *Sensitivity of a modality alone* refers to the number of cancers that would have been detected if only that modality had been used and includes some cancers that were also visible on the other modality.

To be eligible for the MRI substudy, women had to have completed the third round of annual ultrasound and mammography screenings per protocol⁷ and had agreed to undergo contrast-enhanced breast MRI within 8 weeks of the 24-month screening mammogram. Interpretation of each of the 3 screening approaches was blinded to results of the other examinations. A separate integrated breast-level interpretation across all 3 modalities was then performed, which determined clinical management. Women who accepted MRI had higher risk and were younger than those who declined.¹⁵ Women enrolled at sites in the MRI substudy were less likely to have had a personal history of breast cancer; no other systematic differences were noted across sites.

Web-based data capture and quality monitoring were conducted by the ACRIN biostatistics and data management center. The study was compliant with the Health Insurance Portability and Accountability Act, received institutional review board approval from all participating sites and from ACRIN, and received approval from the National Cancer Institute Cancer Imaging Program. The study underwent data and safety monitoring committee review every 6 months.

Participants

Among the 21 sites, 2809 women were recruited between April 2004 and February 2006, 2725 of whom were eligible (FIGURE 1). Women aged at least

25 years presenting for routine mammography were eligible to participate if they met study definitions of elevated risk (Table 1) and had heterogeneously dense or extremely dense parenchyma¹⁴ in at least 1 quadrant, either by prior mammography report or review of prior mammograms. Women were excluded if they were pregnant or lactating or if they had known metastatic disease, signs or symptoms of breast disease, breast surgery within prior 12 months, or breast implants.

For the MRI substudy, women also could not have contraindications to MRI (have a pacemaker, aneurysm clip, or other metallic implant; weigh >135 kg; or have renal impairment [have a glomerular filtration rate of <30 mL/min per 1.73 m² or were undergoing a dialysis regimen]). Participants provided written informed consent at their initial visit. Those participating in the MRI screening provided a second consent at MRI registration.

Screening methods are detailed in the eAppendix (available at <http://www.jama.com>). The expanded 7-point Breast Imaging-Reporting and Data System (BI-RADS)^{14,16,17} assessment scale was used: a score of 1 is negative; 2, benign; 3, probably benign; 4a, low suspicion; 4b, intermediate suspicion; 4c, moderate suspicion; and 5, highly suggestive of malignancy.

Reference Standard

We defined the reference standard, which could be cancer or not, to be the most severe of biopsy results within 365 days of mammographic screening, clinical follow-up at 1 year, or both. Each mammographic screening was targeted for 365 days after the previous mammographic screening. A complete examination of all study breasts performed more than 11 full months after the previous screen was considered the next annual screen; only 88 of 7473 visits (1.2%) occurred before 11 months. The absence of a known diagnosis of cancer for a participant report at interview, review of medical records, or both at least 11 full months (330 days) after mammographic

screening was considered disease negative, as were 7 cases of prophylactic mastectomies with no evidence of cancer at pathology. Biopsy results showing breast cancer (in situ or infiltrating ductal carcinoma or infiltrating lobular carcinoma) in the breast or axillary lymph nodes were considered disease positive.

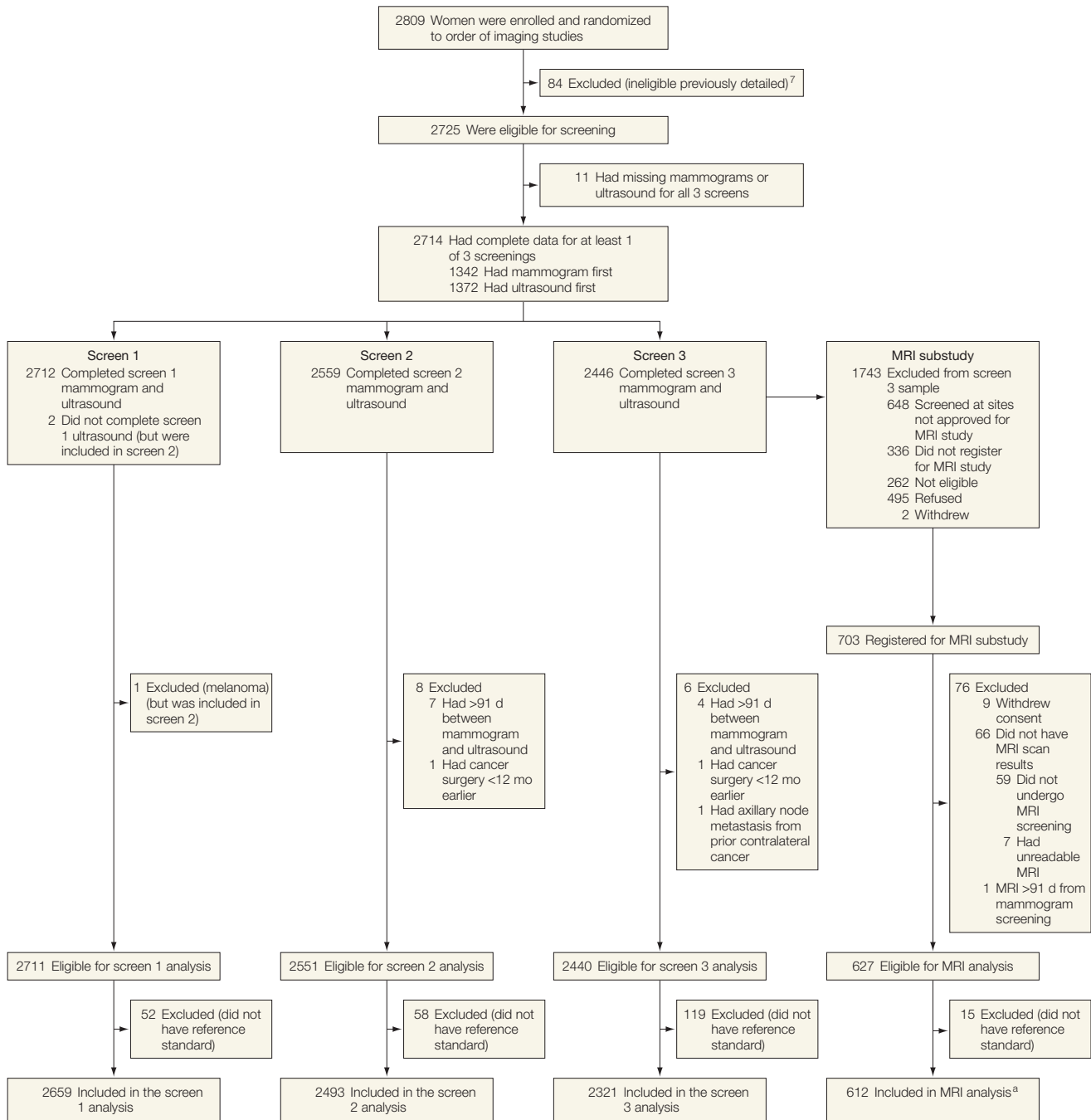
Statistical Methods

The primary unit of analysis was the participant. A participant's BI-RADS score was derived as the maximum breast level BI-RADS or the score from the breast with cancer when only 1 breast had cancer. In keeping with planned revisions to BI-RADS (Edward A. Sickles, MD, Professor of Radiology, University of California, San Francisco, written communication, November 29, 2009), a screening BI-RADS assessment score of 3, 4a, 4b, 4c, or 5 was considered test positive, provided that the recommendation was for other than routine screening. This differs from the definition of positive test results that we used in our initial publication of the first screening, wherein an assessment of 4a or higher was considered a positive test result⁷; results of the first screen have been reanalyzed and included herein. For a participant diagnosed with cancer, the breast(s) with cancer were excluded from analysis for the next annual screen.

The cancer detection rate (ie, the proportion of women with a positive screen result and a positive reference standard); sensitivity; specificity; recall rate, which is the proportion of women with a positive screen result; positive predictive value (PPV1), which is the malignancy rate among cases that test positive on screening; short-term follow-up rate; biopsy rate; and area under the empirical receiver operating characteristic (ROC) curve (AUC) using BI-RADS scores were reported. PPV3 is defined as the rate of malignancy among cases with positive results on screening who underwent biopsy of the same lesion.¹⁴ Interval cancers were defined as those diagnosed because of a clinical abnormality such as a lump, skin

thickening, or pathologic nipple discharge occurring in the interval between prescribed screenings (ie, less than 365 days after the last screening mammogram and before the next screen; cancers detected on an early screen performed at least 11 months after the previous screen were considered screen detected).

Figure 1. Flowchart of Protocol



Participants with negative results on both mammography and ultrasound were imputed as having negative results on integrated reading of mammography plus ultrasound: 1844 for the first screening, 1922 for the second screening, and 1912 for the third screening. The reference standard was the most severe of biopsy results within 365 days of mammographic screening, on clinical follow-up at 1 year, or both. Biopsies prompted by an early subsequent screening examination were attributed to that subsequent screen.

^aAll participants in the magnetic resonance imaging (MRI) analysis set are also in the screen 3 analysis set.

Single-year estimates of yield, sensitivity, specificity, recall rate, PPV1, short-term follow-up rate, biopsy rate, and PPV3, were determined as simple proportions with exact 95% CIs (Clopper-Pearson). The 95% CIs for differences in yield, sensitivity, specificity, recall rate, short-term follow-up rate, and biopsy rate were calculated per Fleiss et al.¹⁸ *P* values for the above comparisons were based on the McNemar test statistic. The 95% CIs and *P* values for differences in PPV1 and PPV3 were calculated using the bootstrap-resampling method.¹⁹ All inferences for incidence screens were based on the bootstrap-resampling method. Estimates, 95% CIs, and *P* values related to AUC were derived by using the method of DeLong et al.²⁰ for empirical ROC curves. Results for participants with a personal history of breast cancer were compared with those who had no such history by the bootstrap method. All *P* values were reported as 2-sided, with .05 set as threshold for significance. All analyses were performed by SAS 9.2 statistical software (SAS Institute Inc).

RESULTS

Participant Demographic Information

A total of 2659 eligible women with reference standard completed the first annual mammogram and ultrasound screenings; 2493, the second; and 2321, the third (Figure 1 and FIGURE 2, Table 1). Participant demographics at enrollment were previously reported.⁷ Median age at enrollment was 55 years (range, 25-91 years). Approximately 29% of women were younger than 50 years at enrollment, and 23% were premenopausal (Table 1). Nearly 54% of women had a personal history of breast cancer. The median age of the 612 women in the MRI group was 57 years (range, 27-87 years); 21% were younger than 50 years at the time of the screening, 25% were premenopausal, and 45% had personal history of breast cancer. Time between screens (eTable 1) and time to perform ultrasound (eTable 2) are available at <http://www.jama.com>.

Cancer Detection

A total of 110 participants were diagnosed with breast cancer during the 3-year study. One woman diagnosed by mammography in the first year was diagnosed again in the third year in the contralateral breast by MRI only. Each diagnosis was counted as a separate event, for a total of 111 participant-cancer events. Of 111 diagnoses, 89 (80%) were invasive (TABLE 2). Fifty-nine cancers (53%) were detected by mammography, including 33 (30%) that were detected by mammography only; 32 (29%) by ultrasound only; and 9 (8%) by MRI only after both mammography and ultrasound screens failed to detect cancer. Eleven cancers (10%) were not detected by any imaging screen. Of 32 cancers seen only on ultrasound, 30 (94%) were invasive, with median size of 10 mm (range, 2-40 mm), and 26 of 27 (96%) of those staged were node negative.

A total of 16 of 612 women (2.6%) in the MRI substudy were diagnosed with breast cancer, 12 of 16 (75%) of whom had invasive cancer. Nine of 16 cancers (56%) were seen only on MRI after negative mammography and ultrasound results: 8 of 9 (89%) were invasive, with median size of 8.5 mm (range, 1-25 mm), and all 7 cancers that were staged were node negative (Table 2). Two invasive cancers that had been detected by ultrasound but not by mammography in the MRI substudy were also detected by MRI.

Supplemental Cancer Detection Yield

Supplemental ultrasound increased cancer detection with each annual screen beyond that of mammography, adding detection of 5.3 cancers per 1000 women in the first year (95% CI, 2.1-8.4; *P* < .001); 3.7 women per 1000 per year in each of the second and third years (95% CI, 2.1-5.8, *P* < .001; TABLE 3); and averaging 4.3 per 1000 for each of the 3 rounds of annual screening. Supplemental yield results of ultrasound after digital mammography are shown in the eAppendix. The addition of MRI screening further in-

creased cancer detection with a supplemental cancer detection yield of 14.7 per 1000 women (95% CI, 3.5-25.9; *P* = .004 vs mammogram plus ultrasound; TABLE 4). The number of screens needed to detect 1 cancer was 127 (95% CI, 99-167) for mammography; 234 (95% CI, 173-345) for supplemental ultrasound, and 68 (95% CI, 39-286) for supplemental MRI after negative mammography plus ultrasound screening results.

Sensitivity, Specificity, and AUC

Among 4814 incidence screens in years 2 and 3 combined, 75 women were diagnosed with cancer. Sensitivity of combined mammography plus ultrasound was 57 of 75 (0.76; 95% CI, 0.65-0.85) for incidence screening, higher than mammography alone, which was 39 of 75 (0.52; 95% CI, 0.40-0.64; *P* < .001). Specificity of combined mammography and ultrasound was 3987 of 4739 (0.84; 95% CI, 0.83 to 0.85) for incidence screens, lower than the specificity of mammography alone, which was 4325 of 4739 (0.91; 95% CI, 0.90-0.92; *P* < .001; Table 3).

For 612 MRI participants, sensitivity increased from 7 of 16 (0.44; 95% CI, 0.20-0.70) with combined mammography and ultrasound to 16 of 16 (1.00; 95% CI, 0.79-1.00) with the addition of MRI (*P* = .004). Specificity was reduced to 390 of 596 (0.65; 95% CI, 0.61-0.69) after MRI vs combined mammography plus ultrasound at 503 of 596 (0.84; 95% CI, 0.81-0.87, *P* < .001; Table 4).

Overall AUC increased each year when ultrasound was added to mammography (Table 3). Adding MRI lowered apparent performance of mammography plus ultrasound because more cancers were identified by MRI (Table 4).

Additional Biopsies and PPV3

The PPV3 for biopsies resulting from combined mammography plus ultrasound was 31 of 272 (0.11; 95% CI, 0.08-0.16) for the first screen and was 55 of 339 (0.16; 95% CI, 0.12 to 0.21) for incidence screens. These values were

significantly lower than those of mammography alone (19 of 65 [0.29; 95% CI, 0.19-0.42, first screening] and 37 of 97 [0.38; 95% CI, 0.28-0.49 incidence screening]; $P < .001$ for both; Table 3). The percentage of women undergoing biopsy after mammography and ultrasound decreased from 272 of 2659 (10.2%; 95% CI, 9.1%-11.4%) in year 1 to 339 of 4814 (7.0%; 95% CI, 6.3%-7.8%) for incidence screens ($P < .001$). The biopsy rates after mammography alone were 65 of 2659 (2.4%; 95% CI, 1.9%-3.1%) in year 1 and 97 of 4814 (2.0%; 95% CI, 1.6%-2.5%) for incidence screens. There were 242 of 4814 (5%) incidence screens resulting in biopsy due to addition of ultrasound, with 18 of 242 (7.4%) of these women found to have cancer.

For 612 MRI participants, the rate of biopsy after a full workup of mammography plus ultrasound was 38 of 612 (6.2%; 95% CI, 4.4%-8.4%), which increased to 81 of 612 (13.2%; 95% CI, 10.7%-16.2%) with the addition of MRI ($P < .001$). The PPV3 after mammography plus ultrasound was 7 of 38 (0.18; 95% CI, 0.08-0.34) and with addition of MRI, it was 15 of 81 (0.19; 95% CI, 0.11-0.29, $P = .98$; Table 4). There were 43 of 612 (7.0%) participants biopsied only because of MRI, 8 (19%) of whom were found to have cancer.

Interval Cancers

Of 20 women with cancer not seen on either mammography or ultrasound in 3 annual rounds, 9 women in the MRI cohort had their cancer detected by MRI. Another 9 cancers were identified because of clinical abnormalities found during the intervals between screens (interval cancer rate 8.1%): 2 had clinical findings in the first year; 4 in the second year; and 3 in the third year. One participant was found to have high-grade ductal carcinoma in situ because of off-study computer-assisted detection applied to mammogram (revealing calcifications) after the year-3 interpretation had been recorded. One participant with a BRCA1 mutation had an MRI screening off study 6 months after the third screen and was found to

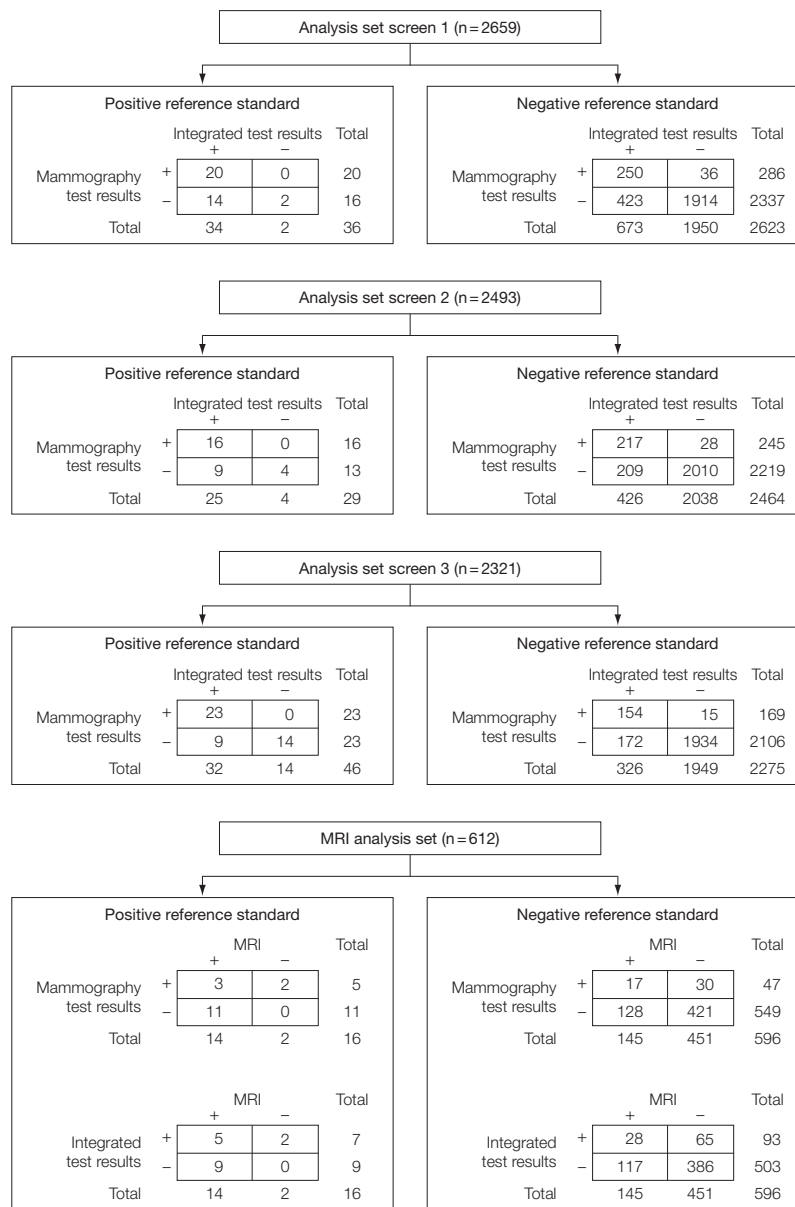
have a 7 mm node-negative grade III invasive ductal carcinoma.

Women With Personal History of Breast Cancer

A total of 1426 of 2659 participants (54%) had a personal history of breast cancer at study entry and underwent 4010 screens; 59 of 1426 (4.1%) were

diagnosed with cancer (28 only ipsilateral and 29 only contralateral to the original cancer; 2 bilateral). Supplemental yield of ultrasound was the same in women with a personal history of breast cancer as in women without a personal history of breast cancer (eTable 3A available at <http://www.jama.com>), as was the absolute increase in

Figure 2. Outcomes of 3 Rounds of Annual Screening Mammography Plus Ultrasound



Outcomes of screening 2662 participants are detailed for mammography alone compared with integrated tests, mammography plus ultrasound, for each of the 3 screening years and also for 612 women in the MRI substudy compared with mammography alone or compared with integrated tests, mammography plus ultrasound, in year 3.

sensitivity due to added ultrasound. Supplemental ultrasound was less likely to prompt unnecessary recall or biopsy in women with a personal history of breast cancer than those without (eTable 3A). The supplemental yield of MRI screening in women with or without a personal history of breast cancer in the MRI substudy is detailed (eTable 3B). The supplemental MRI was less likely to prompt unnecessary recall or biopsy in women with a personal history of breast cancer than those without (eTable 3B).

COMMENT

In this study, annual supplemental incidence screening ultrasound detected an additional 3.7 cancers per 1000 women per year screened beyond mammography alone. The majority of cancers seen only on ultrasound were node-negative invasive

cancers. Invasive lobular carcinoma and low-grade invasive ductal carcinoma were overrepresented among such cancers.

One of the major concerns about screening is the harm of extra testing and biopsies for women who do not have cancer.²¹ As has been observed with mammography²² and MRI,^{11,23-25} the risk of false positives decreased significantly with annual screening ultrasound in this study compared with the first screen. However, there still remained a substantial rate of biopsies prompted only by incidence screening ultrasound, averaging 5.0% of women screened.

In a separate analysis of ACRIN 6666 participants, MRI was significantly less tolerable than mammography or ultrasound. Only 58% of ACRIN 6666 participants who were offered a screening MRI at no out-of-pocket cost accepted

the invitation.¹⁵ These barriers are in addition to high costs of MRI equipment, contrast, and examination, as well as the high rates of induced testing including biopsy, with 7% of women in this study biopsied only because of MRI findings.

Contrast-enhanced MRI has been recommended for supplemental screening of women at high risk of breast cancer, defined as those women with a lifetime risk of 20% to 25% or greater based on family history, known or suspected BRCA or other high-risk genetic mutations, or prior mantle radiation to the chest.²⁶ Across 9 series, the supplemental yield of MRI after mammography in high-risk women was 11 per 1000²⁷ and was 14 per 1000 among the subset who also had screening ultrasound.^{11-13,25} Similar results were observed in this study of women who were mostly at intermediate risk of breast cancer.

Table 2. Summary of Cancer Detection and Characteristics for 2662 Unique Participants Screened 3 Years With Mammography and Physician-Performed Ultrasound and 612 Participants Screened With MRI in Year 3

| | Detected Cancer | | | Not Detected on Study Imaging | Detected by Study MRI Only | Total |
|---|------------------|---------------------------------|----------------------------|-------------------------------|----------------------------|-------------|
| | Mammography Only | Both Mammography and Ultrasound | Ultrasound Only Before MRI | | | |
| No. of participants | 2662 | 2662 | 2662 | NA | 612 | NA |
| No. of screens | 7473 | 7473 | 7473 | NA | 612 | NA |
| No. of cancers | 33 | 26 | 32 | 11 | 9 | 111 |
| Invasive cancers | 18 (55) | 23 (88) | 30 (94) | 10 (91) | 8 (89) | 89 (80) |
| Size invasive tumor, median (range), mm | 11.5 (1-55) | 16.0 (3-40) | 10.0 (2-40) | 8.5 (2-13) | 8.5 (1-25) | 12.0 (1-55) |
| Nodal staging available ^a | 15 | 15 | 27 | 6 | 7 | 70 |
| Node positive, No. (%) | 5 (33) | 7 (47) | 1 (4) | 0 (0) | 0 (0) | 13 (19) |
| Cancer type and grade, No. (%) | | | | | | |
| IDC grade | 17 (52) | 16 (62) | 24 (75) | 8 (73) | 7 (78) | 72 (65) |
| High | 7 (21) | 4 (15) | 6 (19) | 2 (18) | 2 (22) ^b | 21 (19) |
| Intermediate | 6 (18) | 8 (31) | 7 (22) | 1 (9) | 1 (11) | 23 (21) |
| Low | 3 (9) | 4 (15) | 11 (34) | 3 (27) | 4 (44) | 25 (23) |
| Cannot be assessed | 1 (3) | 0 | 0 | 2 (18) | 0 | 3 (3) |
| ILC | 1 (3) | 5 (19) ^c | 5 (16) | 1 (9) | 0 | 12 (11) |
| Mixed IDC and ILC | 0 | 2 (8) ^d | 1 (3) ^d | 1 (9) | 1 (11) | 5 (5) |
| DCIS, nuclear grade | 15 (45) | 3 (12) | 2 (6) | 1 (9) | 1 (11) | 22 (20) |
| High | 2 (6) | 0 | 1 (3) | 1 (9) | 0 | 4 (4) |
| Intermediate | 11 (33) | 3 (12) | 1 (3) | 0 | 0 | 15 (14) |
| Low | 2 (6) | 0 | 0 | 0 | 1 (11) | 3 (3) |

Abbreviations: DCIS, ductal carcinoma in situ; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; NA, not applicable. Grade was collected only for IDC and DCIS.
^aAxillary nodal status could not be assessed for 14 participants with a personal history of breast cancer from whom nodes had previously been removed nor could they be assessed for 1 woman with a personal history of Hodgkin disease and prior nodal treatment. Node status was not determined for 1 participant older than 80 years because it would not affect her treatment planning. For 3 participants without nodal staging, the reason was unknown.
^bIncludes 1 T1mi tumor, with the grade based on the DCIS grade.
^cIncludes 1 ILC with DCIS for which grade of the ILC is missing.
^dIncludes 1 mixed IDC-ILC with associated intermediate nuclear grade DCIS.

Fewer studies have evaluated MRI in women at intermediate risk, including women with a personal history of breast cancer, prior atypical biopsy or lobular carcinoma in situ, intermediate family history of breast cancer (lifetime risk of 15%-20% per the American Cancer Society guidelines²⁶), or women whose only risk factor is dense breasts. Recent studies collectively suggest that supplemental MRI screening may be reasonable for women with a personal history of breast cancer and also found false positives to be less frequent than for women with a family history of breast cancer.²⁸⁻³⁰

For high-risk women unable to undergo MRI,¹⁵ and for intermediate-risk women with dense breasts, including those with a personal history of breast cancer, this study supports the use of supplemental screening with ultrasound in addition to mammography. With either MRI or ultrasound, the risks of false positives, including unnecessary biopsies, were lower for supplemental screening in women with a personal history of breast cancer than in women without. The outcomes in terms of staging, node-positive disease, and interval cancer rates achieved in this study after

3 years of programmatic screening with both ultrasound and mammography were comparable with benchmarks from studies that included MRI.^{10-13,25}

If screening ultrasound were to be adopted for women with dense breasts who are not candidates for MRI, there would be obstacles to its implementation. These include the availability of only 1 current procedural terminology (CPT) code, 76645, for breast ultrasound, with low reimbursement (2010 Medicare reimbursement averaged a global fee of \$89.85 to \$91.83,³¹ which does not cover the costs of physicians performing and

Table 3. Screening Performance in 2662 Unique Participants Screened 3 Years With Mammography and Physician-Performed Ultrasound

| | Mammography Alone | | Combined Mammography Plus Ultrasound | | Difference of (Mammography Plus Ultrasound) and Mammography Alone | | Ultrasound Alone | |
|------------------------------|--------------------|---------------------|--------------------------------------|---------------------|---|---------|--------------------|---------------------|
| | No./Total of Women | Estimate (95% CI) | No./Total of Women | Estimate (95% CI) | Estimate (95% CI) | P Value | No./Total of Women | Estimate (95% CI) |
| Yield, per 1000 | | | | | | | | |
| Screen 1 | 20/2659 | 7.5 (4.6 to 11.6) | 34/2659 | 12.8 (8.9 to 17.8) | 5.3 (2.1 to 8.4) | <.001 | 24/2659 | 9.0 (5.8 to 13.4) |
| Screen 2,3 ^a | 39/4814 | 8.1 (5.8 to 11.1) | 57/4814 | 11.8 (9.0 to 15.3) | 3.7 (2.1 to 5.8) | <.001 | 34/4814 | 7.1 (4.9 to 9.9) |
| AUC | | | | | | | | |
| Screen 1 | | 0.74 (0.63 to 0.84) | | 0.94 (0.89 to 0.99) | 0.20 (0.10 to 0.30) | <.001 | | 0.76 (0.66 to 0.87) |
| Screen 2 | | 0.75 (0.65 to 0.86) | | 0.89 (0.82 to 0.97) | 0.14 (0.03 to 0.25) | .01 | | 0.71 (0.58 to 0.84) |
| Screen 3 | | 0.72 (0.64 to 0.81) | | 0.82 (0.74 to 0.89) | 0.10 (0.00 to 0.18) | .04 | | 0.62 (0.52 to 0.72) |
| Sensitivity, % | | | | | | | | |
| Screen 1 | 20/36 | 55.6 (38.1 to 72.1) | 34/36 | 94.4 (81.3 to 99.3) | 38.9 (20.2 to 57.6) | <.001 | 24/36 | 66.7 (49.0 to 81.4) |
| Screen 2,3 | 39/75 | 52.0 (40.2 to 63.7) | 57/75 | 76.0 (64.7 to 85.1) | 24.0 (14.7 to 33.3) | <.001 | 34/75 | 45.3 (33.8 to 57.3) |
| Specificity, % | | | | | | | | |
| Screen 1 | 2337/2623 | 89.1 (87.8 to 90.3) | 1950/2623 | 74.3 (72.6 to 76.0) | -14.8 (-16.3 to -13.2) | <.001 | 2092/2623 | 79.8 (78.2 to 81.3) |
| Screen 2,3 | 4325/4739 | 91.3 (90.4 to 92.1) | 3987/4739 | 84.1 (83.1 to 85.2) | -7.1 (-8.0 to -6.3) | <.001 | 4258/4739 | 89.9 (89.0 to 90.7) |
| Recall rate, % | | | | | | | | |
| Screen 1 | 306/2659 | 11.5 (10.3 to 12.8) | 707/2659 | 26.6 (24.9 to 28.3) | 15.1 (13.5 to 16.6) | <.001 | 555/2659 | 20.9 (19.3 to 22.5) |
| Screen 2,3 | 453/4814 | 9.4 (8.6 to 10.3) | 809/4814 | 16.8 (15.8 to 17.9) | 7.4 (6.6 to 8.2) | <.001 | 515/4814 | 10.7 (9.8 to 11.6) |
| PPV1, % ^b | | | | | | | | |
| Screen 1 | 20/306 | 6.5 (4.0 to 9.9) | 34/707 | 4.8 (3.4 to 6.7) | -1.7 (-3.7 to 0.1) | .07 | 24/555 | 4.3 (2.8 to 6.4) |
| Screen 2,3 | 39/453 | 8.6 (6.2 to 11.6) | 57/809 | 7.0 (5.4 to 9.0) | -1.6 (-3.1 to -0.2) | .04 | 34/515 | 6.6 (4.6 to 9.1) |
| Short-term follow-up rate, % | | | | | | | | |
| Screen 1 | 84/2659 | 3.2 (2.5 to 3.9) | 368/2659 | 13.8 (12.5 to 15.2) | 10.7 (9.5 to 11.9) | <.001 | 296/2659 | 11.1 (10.0 to 12.4) |
| Screen 2,3 | 76/4814 | 1.6 (1.2 to 2.0) | 256/4814 | 5.3 (4.7 to 6.0) | 3.7 (3.2 to 4.3) | <.001 | 190/4814 | 3.9 (3.4 to 4.5) |
| Biopsy rate, % | | | | | | | | |
| Screen 1 | 65/2659 | 2.4 (1.9 to 3.1) | 272/2659 | 10.2 (9.1 to 11.4) | 7.8 (6.7 to 8.8) | <.001 | 233/2659 | 8.8 (7.7 to 9.9) |
| Screen 2,3 | 97/4814 | 2.0 (1.6 to 2.5) | 339/4814 | 7.0 (6.3 to 7.8) | 5.0 (4.4 to 5.7) | <.001 | 266/4814 | 5.5 (4.9 to 6.2) |
| PPV3, % ^c | | | | | | | | |
| Screen 1 | 19/65 | 29.2 (18.6 to 41.8) | 31/272 | 11.4 (7.9 to 15.8) | -17.8 (-26.7 to -9.3) | <.001 | 21/233 | 9.0 (5.7 to 13.4) |
| Screen 2,3 | 37/97 | 38.1 (28.5 to 48.6) | 55/339 | 16.2 (12.5 to 20.6) | -21.9 (-28.7 to -14.7) | <.001 | 31/266 | 11.7 (8.1 to 16.1) |

Abbreviations: AUC, area under the curve; PPV, positive predictive value.

^aScreen 2,3 refers to incidence screens in years 2 and 3 (ie, at 12 and 24 mo after study entry respectively).

^bDefined as the malignancy rate among women with a positive screening test (ie, assessment of BI-RADS 3 or higher and recalled from screening for further testing or short-interval follow-up).

^cDefined as the malignancy rate among women with a positive screening test who underwent biopsy of the same lesion.

interpreting a thorough screening examination). While supplemental cancer detection rates with technologist-performed screening ultrasound were similar to physician-performed ultrasound in one series,⁴ there remains a shortage of qualified breast ultrasound technologists.

There are a few limitations to this study. Additional node-negative invasive cancers were found by adding screening ultrasound to mammography in each incidence screen, and increasing detection of such cancers correlates with mortality reduction.³² However, we did not have a control group with no ultrasound performed

with which we could compare clinical outcomes, and mortality was not assessed. In Japan, the ongoing Japan Strategic Anti-Cancer Randomized Trial (J-START) of biennial mammography, with or without technologist-performed screening ultrasound does have such a control group.³³ We only performed a single screening MRI, and false positives would be expected to decrease in subsequent years.^{11,23} Not all sites in the original ACRIN 6666 protocol were able to offer MRI.

CONCLUSION

The cancer detection benefit from supplemental screening ultrasound seen

on the first screening persisted with each annual screening. Rates of biopsy for findings seen only on ultrasound remained substantial on incidence screens, representing 5% of women, with only 7.4% of those women found to have cancer. Risks of false-positives were lower in women with a personal history of breast cancer than in women without.

As has been seen in other studies,^{10,11,25,34} MRI significantly increased detection of early breast cancer beyond that seen with mammography or mammography combined with ultrasound. The 56% absolute increase in cancer detection seen in the MRI substudy (16 of

Table 4. Screening Performance in 612 Participants Screened by Magnetic Resonance Imaging After 3 Annual Mammography and Ultrasound Screenings

| | Combined Mammography Plus Ultrasound | Combined Mammography Plus Ultrasound Plus MRI | Difference of (Mammography Plus Ultrasound Plus MRI) and (Mammography Plus Ultrasound) | | Mammography Alone | Combined Mammography Plus MRI | Difference of (Mammography Plus MRI) and Mammography Alone | | MRI Alone |
|---------------------------------------|--------------------------------------|---|--|----------------------|---------------------|-------------------------------|--|----------------------|---------------------|
| | | | Estimate (95% CI) | P Value ^a | | | Estimate (95% CI) | P Value ^b | |
| Yield (95% CI), per 1000 ^c | 11.4 (4.6 to 23.4) | 26.1 (15.0 to 42.1) | 14.7 (3.5 to 25.9) | .004 | 8.2 (2.7 to 19.0) | 26.1 (15.0 to 42.1) | 18.0 (5.8 to 30.1) | <.001 | 22.9 (12.6 to 38.1) |
| No./total | 7/612 | 16/612 | | | 5/612 | 16/612 | | | 14/612 |
| AUC (95% CI) | 0.69 (0.55 to 0.83) | 0.95 (0.91 to 0.99) | 0.26 (0.11 to 0.42) | <.001 | 0.63 (0.47 to 0.78) | 0.94 (0.90 to 0.98) | 0.31 (0.16 to 0.46) | <.001 | 0.87 (0.75 to 0.98) |
| Sensitivity (95% CI), % | 43.8 (19.8 to 70.1) | 100.0 (79.4 to 100.0) | 56.3 (25.7 to 86.8) | .004 | 31.3 (11.0 to 58.7) | 100.0 (79.4 to 100.0) | 68.8 (39.8 to 97.7) | <.001 | 87.5 (61.7 to 98.4) |
| No./total | 7/16 | 16/16 | | | 5/16 | 16/16 | | | 14/16 |
| Specificity (95% CI), % | 84.4 (81.2 to 87.2) | 65.4 (61.5 to 69.3) | -19.0 (-22.3 to -15.6) | <.001 | 92.1 (89.7 to 94.1) | 70.6 (66.8 to 74.3) | -21.5 (-24.9 to -18.0) | <.001 | 75.7 (72.0 to 79.1) |
| No./total | 503/596 | 390/596 | | | 549/596 | 421/596 | | | 451/596 |
| Recall rate (95% CI), % | 16.3 (13.5 to 19.5) | 36.3 (32.5 to 40.2) | 19.9 (16.6 to 23.3) | <.001 | 8.5 (6.4 to 11.0) | 31.2 (27.6 to 35.0) | 22.7 (19.2 to 26.2) | <.001 | 26.0 (22.5 to 29.6) |
| No./total | 100/612 | 222/612 | | | 52/612 | 191/612 | | | 159/612 |
| PPV1 (95% CI), % ^d | 7.0 (2.9 to 13.9) | 7.2 (4.2 to 11.4) | 0.2 (-3.8 to 4.0) | .92 | 9.6 (3.2 to 21.0) | 8.4 (4.9 to 13.2) | -1.2 (-8.0 to 4.6) | .70 | 8.8 (4.9 to 14.3) |
| No./total | 7/100 | 16/222 | | | 5/52 | 16/191 | | | 14/159 |
| Short-term follow-up rate (95% CI), % | 4.6 (3.1 to 6.5) | 19.6 (16.5 to 23.0) | 15.0 (12.0 to 18.0) | <.001 | 0.5 (0.1 to 1.4) | 16.3 (13.5 to 19.5) | 15.8 (12.8 to 18.9) | <.001 | 15.8 (13.0 to 19.0) |
| No./total | 28/612 | 120/612 | | | 3/612 | 100/612 | | | 97/612 |
| Biopsy rate (95% CI), % | 6.2 (4.4 to 8.4) | 13.2 (10.7 to 16.2) | 7.0 (4.8 to 9.2) | <.001 | 1.6 (0.8 to 3.0) | 9.6 (7.4 to 12.3) | 8.0 (5.7 to 10.3) | <.001 | 8.5 (6.4 to 11.0) |
| No./total | 38/612 | 81/612 | | | 10/612 | 59/612 | | | 52/612 |
| PPV3 (95% CI), % ^e | 18.4 (7.7 to 34.3) | 18.5 (10.8 to 28.7) | 0.1 (-8.8 to 8.8) | .98 | 50.0 (18.7 to 81.3) | 25.4 (15.0 to 38.4) | -24.6 (-51.2 to 3.7) | .08 | 23.1 (12.5 to 36.8) |
| No./total | 7/38 | 15/81 | | | 5/10 | 15/59 | | | 12/52 |

Abbreviation: MRI, magnetic resonance imaging; PPV, positive predictive value.
^aP value that observed difference of combined mammography plus ultrasound, and MRI vs mammography plus ultrasound occurred by chance.
^bP value that observed difference of combined mammography and MRI vs mammography alone occurred by chance.
^cYield is the cancer detection rate.
^dDefined as the malignancy rate among women with a positive screening test (ie, assessment of BI-RADS 3 or higher and recalled from screening for further testing or short-interval follow-up).
^eDefined as the malignancy rate among women with a positive screening test who underwent biopsy of the same lesion.

16 vs 7 of 16) was greater than the 34% absolute increase in invasive cancer detection (71 of 89 vs 41 of 89) seen by adding annual ultrasound to mammography in the main ACRIN 6666 study. However, given the low clinically detected interval cancer rate of 8% in the main ACRIN 6666 protocol and given the fact that all interval cancers remained node-negative at diagnosis, it is unclear that the added cost and reduced tolerability of screening MRI are justified in women at intermediate risk for breast cancer in lieu of supplemental screening with ultrasound. Despite its higher sensitivity, the addition of screening MRI rather than ultrasound to mammography in broader populations of women at intermediate risk with dense breasts may not be appropriate, particularly when the current high false-positive rates, cost, and reduced tolerability of MRI are considered.

Author Affiliations: American College of Radiology Imaging Network, Philadelphia, Pennsylvania (Dr Berg and Ms Gabrielli); Center for Statistical Sciences, Brown University, Providence, Rhode Island (Drs Zhang and Cormack and Mss Marques, Adams, and Yeh); CERIM, Buenos Aires, Argentina (Dr Lehrer); Department of Radiology, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada (Dr Jong); Department of Radiology, University of North Carolina, Chapel Hill (Dr Pisano); Radiology Consultants/Forum Health, Youngstown, Ohio (Dr Barr); Weinstein Imaging Associates, Pittsburgh, Pennsylvania (Dr Böhm-Vélez); Department of Radiology, University of Cincinnati Medical Center, Cincinnati, Ohio (Dr Mahoney); Department of Radiology, University of Texas Southwestern Medical Center, Dallas (Dr Evans); Department of Radiology, Keck School of Medicine, University of Southern California, Los Angeles (Dr Larsen); Department of Radiology, Mayo Clinic, Rochester, Minnesota (Dr Morton); Department of Radiology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois (Dr Mendelson); Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, Missouri (Dr Farria). Dr Berg is now with the Department of Radiology, Magee-Womens Hospital, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, and Dr Pisano is now with the Department of Radiology, Medical University of South Carolina, Charleston.

Author Contributions: Drs Zhang and Cormack had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Berg, Zhang, Mendelson, Pisano.

Acquisition of data: Berg, Mendelson, Lehrer, Böhm-Vélez, Pisano, Jong, Evans, Morton, Mahoney, Hovanessian-Larsen, Barr, Farria, Gabrielli.

Analysis and interpretation of data: Berg, Zhang, Cormack, Pisano, Marques, Adams, Yeh.

Drafting of the manuscript: Berg, Zhang.

Critical revision of the manuscript for important intellectual content: Berg, Zhang, Mendelson, Lehrer,

Böhm-Vélez, Pisano, Jong, Evans, Morton, Mahoney, Hovanessian-Larsen, Barr, Farria, Gabrielli.

Statistical analysis: Zhang, Cormack, Marques, Adams, Yeh.

Obtained funding: Berg.

Administrative, technical, or material support: Berg, Mendelson, Gabrielli.

Study supervision: Berg, Zhang.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Berg reports that she has served as a consultant to Naviscan Inc and SuperSonic Imagine, has received research support from Naviscan Inc, has prepared educational materials for Gamma Medica, has a research grant from Hologic Inc, and is on the medical advisory board of Philips. Dr Mendelson reports that she is a member of the scientific advisory boards of MediPattern, Hologic, and Siemens and has received equipment support from Philips and research support from SuperSonic Imagine and Siemens. Dr Böhm-Vélez reports that she is a member of the scientific advisory board of Philips, does clinical validation studies for Philips Ultrasound, and is on the speakers bureau of Dilon. Dr Pisano reports that her laboratory received research support from GE Healthcare, Konica Minolta, Sectra AB, Naviscan Inc, Koning, Zumatek, Inc, equipment grants from R2 and iCAD, is a board member of ACR Imaging Matrix and NextRay Inc, and a stockholder in NextRay Inc. Dr Jong reports that she is a consultant to and receives research support from GE Healthcare. Dr Evans reports that he is a member of the scientific advisory board of Hologic. Dr Mahoney reports that she is a consultant to Ethicon EndoSurgery and SenoRx and on the scientific advisory board of Hologic and receives research support from Naviscan Inc. Dr Larsen reports that she receives equipment support from Naviscan Inc. Dr Barr reports that he is a member of the ultrasound advisory boards of and has received equipment support, research support, and speakers fees from Siemens and Philips, an equipment grant from SuperSonic Inc, and a research grant from Bracco. The remaining coauthors report no financial disclosures.

ACRIN 6666 Site Investigators and Research Assistants: *Allegheny-Singer Research Institute, Pittsburgh, Pennsylvania:* William R. Poller, MD, principal investigator (PI), and Michelle Huerbin; *American Radiology Services—Johns Hopkins Green Spring, Baltimore, Maryland:* Wendie A. Berg, MD, PhD (PI), Barbara E. Levit, RT, and Kathy Wetzell; *Beth Israel Deaconess Medical Center, Boston, Massachusetts:* Janet K. Baum, MD, and Valerie J. Fein-Zachary, MD (PIs), and Suzette M. Kelleher, BA; *CERIM, Buenos Aires, Argentina:* Daniel E. Lehrer, MD (PI), and Maria S. Osertag; *Duke University Medical Center, Durham, North Carolina:* Mary Scott Soo, MD (PI), and Brenda N. Prince, RT; *Mayo Clinic, Rochester, Minnesota:* Marilyn J. Morton, DO (PI), and Lori M. Johnson, AAS; *Feinberg School of Medicine, Northwestern University, Chicago, Illinois:* Ellen B. Mendelson, MD (PI), and Marysia Kalata, AA; *Radiology Associates of Atlanta, Atlanta, Georgia:* Handel Reynolds, MD (PI), and Y. Suzette Wheeler, RN, MSHA; *Radiology Consultants/Forum Health, Youngstown, Ohio:* Richard G. Barr, MD, PhD (PI), and Marilyn J. Mangino, RN; *Radiology Imaging Associates, Denver, Colorado:* A. Thomas Stavros, MD (PI), and Margo Valdez; *Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada:* Roberta A. Jong, MD (PI), and Julie H. Lee, BSc; *Thomas Jefferson University Hospital, Philadelphia, Pennsylvania:* Catherine W. Piccoli, MD, and Christopher R. B. Merritt, MS, MD (PIs), and Colleen Descenzo; *David Geffen School of Medicine at University of California Los Angeles Medical Center, Los Angeles:* Anne C. Hoyt, MD (PI), and Roslynn Marzan, BS; *University of Cincinnati Medical Center, Cincinnati, Ohio:* Mary C. Mahoney, MD (PI), and Monene M. Kamm, AS; *University of North*

Carolina, Chapel Hill: Etta D. Pisano, MD (PI), and Laura A. Tuttle, MA; *Keck School of Medicine, University of Southern California, Los Angeles:* Linda H. Larsen, MD (PI), and Christina E. Kiss, AA; *University of Texas MFD Anderson Cancer Center, Houston:* Gary J. Whitman, MD (PI), and Sharon R. Rice, AA; *University of Texas Southwestern Medical Center, Dallas:* W. Phil Evans, MD (PI), and Kimberly T. Taylor, AA; *Washington University School of Medicine, St Louis, Missouri:* Dione M. Farria, MD, MPH (PI), and Darlene J. Bird, RT, AS; and *Weinstein Imaging Associates, Pittsburgh, Pennsylvania:* Marcela Böhm-Vélez, MD, (PI), and Antoinette Cockroft.

Funding/Support: The study was funded by the Avon Foundation and grants CA 80098 and CA 79778 from the National Cancer Institute.

Role of the Sponsors: The Avon Foundation was not involved in the design or conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of the manuscript. The trial was conducted by the American College of Radiology Imaging Network, a member of the National Cancer Institute's Clinical Trials Cooperative Groups Program, and was developed and carried out adhering to the standard cooperative group processes. These processes include review of and input about the trial design from the NCI's Cancer Therapy Evaluation Program (CTEP). Upon CTEP's approval of the research protocol, the NCI was not involved in the design or conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of the manuscript.

Previous Presentation: Presented at the 2009 Radiologic Society of North America Scientific Assembly, Chicago, Illinois, December 1, 2009.

Online-Only Material: The eAppendices, eTables, and Author Video Interview are available at <http://www.jama.com>. The full protocol is available online at <http://acr.org/Portals/0/Protocols/6666/Protocol-ACRIN%206666%20Admin%20Update%2011.30.07.pdf>.

Additional Contributions: We thank Jeffrey Blume, PhD, Vanderbilt University, Nashville, Tennessee, for his contributions to initial study design and supervision and Robert A. Smith, PhD, American Cancer Society, Atlanta, Georgia, for review and helpful discussions. We especially thank Marydale DeBor, JD, Chief Advisor Avon Breast Cancer Crusade 1993-2003, who was instrumental in securing the support within the Avon Foundation that made this study possible, and Marc Hurlbert, PhD, of the Avon Foundation for his continued vision and unwavering support. We are indebted to the many investigators, coinvestigators, and research assistants at the clinical sites. We appreciate the efforts of ACRIN Data Management and Imaging staff, and we especially thank Cynthia Olson, MBA, MHS, for administrative oversight. We thank Eric Berns, PhD, University of Colorado, Denver, for ultrasound image quality control and R. Edward Hendrick, PhD, University of Colorado, for MRI imaging quality control. No one was compensated beyond their usual salary for their efforts for this study. We are especially grateful to the 2809 women who enrolled in this study.

REFERENCES

- Buchberger W, Niehoff A, Obrist P, DeKoekkoek-Doll P, Dünser M. Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. *Semin Ultrasound CT MR*. 2000; 21(4):325-336.
- Crystal P, Strano SD, Shcharynski S, Koretz MJ. Using sonography to screen women with mammographically dense breasts. *AJR Am J Roentgenol*. 2003; 181(1):177-182.
- Gordon PB, Goldenberg SL. Malignant breast masses detected only by ultrasound. *Cancer*. 1995;76(4):626-630.
- Kaplan SS. Clinical utility of bilateral whole-breast

- US in the evaluation of women with dense breast tissue. *Radiology*. 2001;221(3):641-649.
5. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them. *Radiology*. 2002;225(1):165-175.
 6. Leconte I, Feger C, Galant C, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. *AJR Am J Roentgenol*. 2003;180(6):1675-1679.
 7. Berg WA, Blume JD, Cormack JB, et al; ACRIN 6666 Investigators. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA*. 2008;299(18):2151-2163.
 8. Corsetti V, Ferrari A, Ghirardi M, et al. Role of ultrasonography in detecting mammographically occult breast carcinoma in women with dense breasts. *Radiol Med*. 2006;111(3):440-448.
 9. Tohno E, Ueno E, Watanabe H. Ultrasound screening of breast cancer. *Breast Cancer*. 2009;16(1):18-22.
 10. Kuhl C, Weigel S, Schrading S, et al. Prospective multicenter cohort study to refine management recommendations for women at elevated familial risk of breast cancer. *J Clin Oncol*. 2010;28(9):1450-1457.
 11. Kuhl CK, Schrading S, Leutner CC, et al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *J Clin Oncol*. 2005;23(33):8469-8476.
 12. Lehman CD, Isaacs C, Schnall MD, et al. Cancer yield of mammography, MR, and US in high-risk women. *Radiology*. 2007;244(2):381-388.
 13. Sardaneli F, Podo F, D'Agnolo G, et al; High Breast Cancer Risk Italian Trial. Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRIT study). *Radiology*. 2007;242(3):698-715.
 14. D'Orsi CJ, Bassett LW, Berg WA, et al. *Breast Imaging Reporting and Data System, BI-RADS: Mammography*. 4th ed. Reston, VA: American College of Radiology; 2003.
 15. Berg WA, Blume JD, Adams AM, et al. Reasons women at elevated risk of breast cancer refuse breast MR imaging screening. *Radiology*. 2010;254(1):79-87.
 16. Mendelson EB, Baum JK, Berg WA, Merritt CRB, Rubin E. *Breast Imaging Reporting and Data System, BI-RADS: Ultrasound*. Reston, VA: American College of Radiology; 2003.
 17. Ikeda DM, Hylton NM, Kuhl CK et al. *Breast Imaging Reporting and Data System, BI-RADS: Magnetic Resonance Imaging*. Reston, VA: American College of Radiology; 2003.
 18. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. Hoboken, NJ: Wiley-Interscience; 2003.
 19. Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat*. 1979;7(1):1-26.
 20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves. *Biometrics*. 1988;44(3):837-845.
 21. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer. *Ann Intern Med*. 2009;151(10):727-737.
 22. Schell MJ, Yankaskas BC, Ballard-Barbash R, et al. Evidence-based target recall rates for screening mammography. *Radiology*. 2007;243(3):681-689.
 23. Kriege M, Brekelmans CT, Boetes C, et al; Dutch MRI Screening (MRISC) Study Group. Differences between first and subsequent rounds of the MRISC breast cancer screening program for women with a familial or genetic predisposition. *Cancer*. 2006;106(11):2318-2326.
 24. Leach MO, Boggis CR, Dixon AK, et al; MARIBS study group. Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer. *Lancet*. 2005;365(9473):1769-1778.
 25. Warner E, Plewes DB, Hill KA, et al. Surveillance of *BRCA1* and *BRCA2* mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA*. 2004;292(11):1317-1325.
 26. Saslow D, Boetes C, Burke W, et al; American Cancer Society Breast Cancer Advisory Group. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin*. 2007;57(2):75-89.
 27. Berg WA. Tailored supplemental screening for breast cancer. *AJR Am J Roentgenol*. 2009;192(2):390-399.
 28. Brennan ME, Houssami N, Lord S, et al. Magnetic resonance imaging screening of the contralateral breast in women with newly diagnosed breast cancer. *J Clin Oncol*. 2009;27(33):5640-5649.
 29. Brennan S, Liberman L, Dershaw DD, Morris E. Breast MRI screening of women with a personal history of breast cancer. *AJR Am J Roentgenol*. 2010;195(2):510-516.
 30. Demartini WB, Kalish GM, Peacock S, Eby PR, Gutierrez RL, Lehman CD. Screening MRI for high risk women. Paper presented at: Radiologic Society of North America; November 28, 2010; Chicago, IL.
 31. Centers for Medicare & Medicaid Services. Physician fee schedule search page. <http://www.cms.gov/apps/physician-fee-schedule>. Accessed March 9, 2012.
 32. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AM, Chen TH. The randomized trials of breast cancer screening. *Radiol Clin North Am*. 2004;42(5):793-806.
 33. Ohuchi N, Ishida T, Kawai M, et al. Randomized controlled trial on effectiveness of ultrasonography screening for breast cancer in women aged 40-49 (J-START). *Jpn J Clin Oncol*. 2011;41(2):275-277.
 34. Sardaneli F, Podo F, Santoro F, et al; High Breast Cancer Risk Italian 1 (HIBCRIT-1) Study. Multicenter surveillance of women at high genetic breast cancer risk using mammography, ultrasonography, and contrast-enhanced magnetic resonance imaging (the high breast cancer risk italian 1 study). *Invest Radiol*. 2011;46(2):94-105.

The California Breast Density Information Group: A Collaborative Response to the Issues of Breast Density, Breast Cancer Risk, and Breast Density Notification Legislation¹

Elissa R. Price, MD
Jonathan Hargreaves, MD
Jafi A. Lipson, MD
Edward A. Sickles, MD
R. James Brenner, MD, JD
Karen K. Lindfors, MD, MPH
Bonnie N. Joe, MD, PhD
Jessica W. T. Leung, MD
Stephen A. Feig, MD
Lawrence W. Bassett, MD
Haydee Ojeda-Fournier, MD
Bruce L. Daniel, MD
Allison W. Kurian, MD, MSc
Elyse Love, MS, CGC
Lauren Ryan, MS, LCGC
Donna D. Walgenbach, MS, CGC
Debra M. Ikeda, MD

In anticipation of breast density notification legislation in the state of California, which would require notification of women with heterogeneously and extremely dense breast tissue, a working group of breast imagers and breast cancer risk specialists was formed to provide a common response framework. The California Breast Density Information Group identified key elements and implications of the law, researching scientific evidence needed to develop a robust response. In particular, issues of risk associated with dense breast tissue, masking of cancers by dense tissue on mammograms, and the efficacy, benefits, and harms of supplementary screening tests were studied and consensus reached. National guidelines and peer-reviewed published literature were used to recommend that women with dense breast tissue at screening mammography follow supplemental screening guidelines based on breast cancer risk assessment. The goal of developing educational materials for referring clinicians and patients was reached with the construction of an easily accessible Web site that contains information about breast density, breast cancer risk assessment, and supplementary imaging. This multi-institutional, multidisciplinary approach may be useful for organizations to frame responses as similar legislation is passed across the United States.

© RSNA, 2013

Supplemental material: <http://radiology.rsna.org/lookup/suppl/doi:10.1148/radiol.13131217/-/DC1>

¹From the Departments of Radiology and Biomedical Imaging, Division of Women's Imaging (E.R.P., E.A.S., B.N.J.), and Radiology (R.J.B.), University of California, San Francisco, San Francisco, Calif; Department of Radiology (J.H., K.K.L.) and the Comprehensive Cancer Center (D.D.W.), University of California, Davis, Sacramento, Calif; Department of Radiology, Stanford University School of Medicine, Advanced Medicine Center, 875 Blake Wilbur Dr, Room CC-2239, Stanford, Calif (J.A.L., D.M.I.); Bay Imaging Consultants, Sutter Health, Alta Bates Summit Medical Center, Carol Ann Read Breast Health Center, Oakland, Calif (R.J.B.); Department of Radiology, Sutter Health, California Pacific Medical Center, San Francisco, Calif (J.W.T.L.); Department of Radiology, University of California, Irvine Medical Center, Fong and Jean Tsai Professor of Women's Imaging, University of California Irvine School of Medicine, UCI Medical Center, Orange, Calif (S.A.F.); Department of Radiology, University of California, Los Angeles, Los Angeles, Calif (L.W.B.); Department of Clinical Radiology, Moores Cancer Center, UC San Diego Health System, La Jolla, Calif (H.O.F.); Department of Radiology, Stanford University School of Medicine, Stanford, Calif (B.L.D.); Divisions of Oncology and Epidemiology, Stanford University School of Medicine, Stanford, Calif (A.W.K.); Department of OB/GYN, UC Davis Health System, University of California Davis Cancer Center, Sacramento, Calif (E.L.); and Athena Breast Health Network and UCSF Cancer Risk Program, San Francisco, Calif (L.R.). Received June 6, 2013; revision requested July 19; revision received and final version accepted August 8. Address correspondence to J.A.L. (e-mail: jlipson@stanford.edu).

© RSNA, 2013

In 2009, Connecticut enacted a law mandating patient and referring physician notification when the pattern of fibroglandular tissue on a patient's mammogram was considered dense by the interpreting radiologist. Similar bills have since been proposed in many other states, nine of which have already passed into law. In California, mandatory reporting requirements took effect on April 1, 2013.

Radiologists are now faced with responding to both patients and referring physicians in trying to reconcile the legislative intent of density notification with realistic and practical practice patterns. In California, a working group of breast imagers and breast cancer risk specialists was formed soon after the passage of the law in an attempt to provide a common response framework. The California Breast Density Information Group (CBDIG) was composed of academic and community-based specialists and began a series of weekly conference calls aiming to produce accessible and valuable materials. We recognized that many institutions would also respond individually, based on local concerns, preferences, and available resources. The purpose of the coalition was to leverage the expertise of practitioners at multiple California institutions to develop an evidence-based consensus. This deliberative process could also provide a model that physicians in

other states could use to develop their own response to pending or already enacted legislation.

The key issues involved in breast density notification involve (a) the relative risk of breast cancer associated with dense breasts, (b) masking of cancers by overlying breast tissue at mammography, and (c) the efficacy, benefits, and harms of supplementary screening tests. We sought to provide a balanced viewpoint on the available scientific data, independent of positions advanced by the manufacturing sector, radiology practices with existing scientific or financial investments in specific supplementary screening technology, and patient advocacy groups. Our goal was to construct an online resource of user-friendly, evidence-based information for patients, referring clinicians, and radiologists.

Our work led us to the development of a document suitable for Internet access, entitled "Frequently Asked Questions About Breast Density, Breast Cancer Risk, and the Breast Density Notification Law in California: A Consensus Document." We concluded that an online working document, although easily accessible and navigated, would also allow us to promptly update information as new scientific data become available. The document is available free online at www.breastdensity.info and included in Appendix E1 (online). The legislated notification statements are in Appendix E2 (online).

The remainder of this special report is a summary of our discussions, the issues practitioners are likely to encounter, and a brief review of the most pertinent literature that supports our frequently asked questions document.

CBDIG Position on Breast Density and Its Implications

Breast density is currently classified by the subjective visual assessment of the interpreting physician into one of four categories, as defined by the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) (1) (Figure). Although new technologies provide quantitative

density assessment, these are not part of common practice. On the basis of large-scale population-based data from a representative sample of mammography practices in the United States, the frequency distribution of the BI-RADS density categories is approximately as follows: almost entirely fatty, 10%; scattered areas of fibroglandular density, 40%; heterogeneously dense, 40%; and extremely dense, 10% (2). All women who fall into the heterogeneously dense and extremely dense categories—approximately 50% of women who undergo screening mammography—must be informed that they have dense breasts under the California law. Approximately 4 million screening mammography examinations are performed annually in California (3–6). Therefore, taking into consideration both screening and diagnostic mammographic examinations, more than 2 million women will receive a density notification letter each year in this state alone.

One important effect of increased breast density is a decrease in mammographic sensitivity (masking). It has been demonstrated that mammographic sensitivity is diminished in dense breasts (7–12). The magnitude of this decrease varies depending on patient age, the density categories that

Advances in Knowledge

- Approximately 50% of women have breast tissue that is classified by the interpreting physician as either heterogeneously dense or extremely dense, translating into approximately 2 million women in the state of California receiving a density notification letter annually.
- Primary issues relating to breast density include the cancer risk imparted by dense tissue and the masking effect.
- For patients who are interested in additional screening options, a breast cancer risk assessment may be appropriate.

Published online before print

10.1148/radiol.13131217 Content codes: **BR** **HP**

Radiology 2013; 269:887–892

Abbreviations:

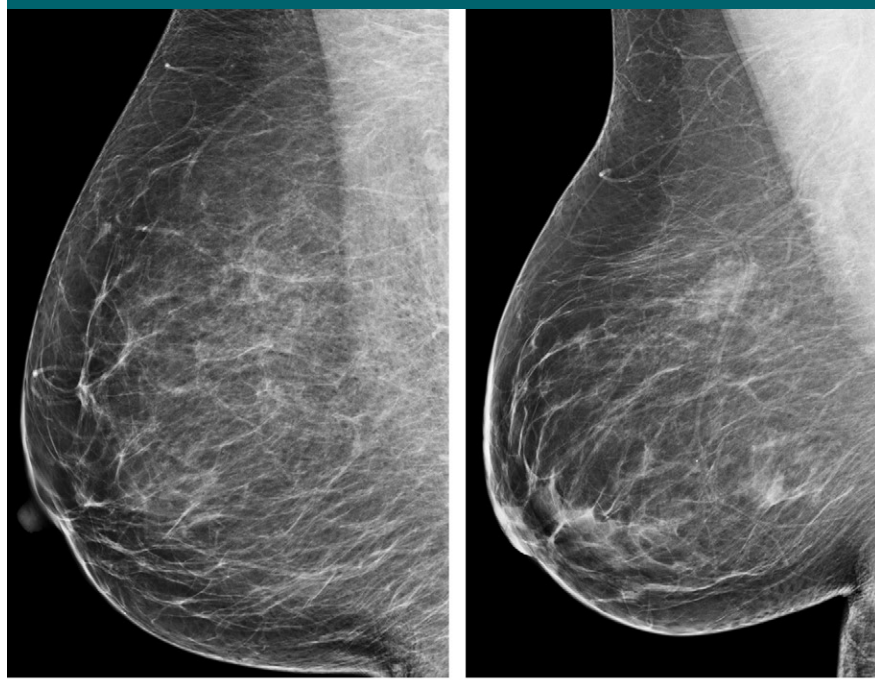
BI-RADS = Breast Imaging Reporting and Data System
 CBDIG = California Breast Density Information Group
 FDA = Food and Drug Administration
 3D = three-dimensional

Author contributions:

Guarantors of integrity of entire study, E.R.P., J.H., J.A.L., B.N.J., L.W.B.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, E.R.P., J.H., J.A.L., E.A.S., R.J.B., K.K.L., B.N.J., S.A.F., A.W.K., D.M.I.; statistical analysis, E.A.S.; and manuscript editing, E.R.P., J.H., J.A.L., E.A.S., R.J.B., K.K.L., B.N.J., J.W.T.L., S.A.F., L.W.B., H.O.F., B.L.D., A.W.K., D.M.I.

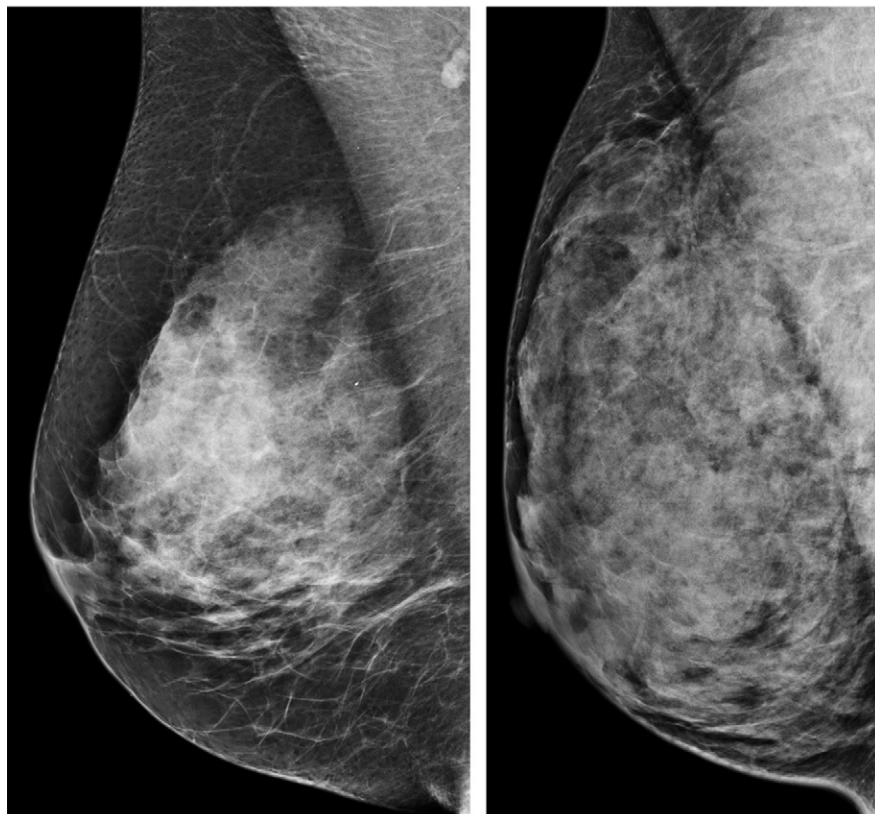
Conflicts of interest are listed at the end of this article.

See also the editorial by D'Orsi in this issue.



a.

b.



c.

d.

Mediolateral oblique mammographic views demonstrate the four BI-RADS breast density categories. **(a)** Almost entirely fatty. **(b)** Scattered fibroglandular density. **(c)** Heterogeneously dense, which may obscure detection of small masses. **(d)** Extremely dense, which lowers the sensitivity of mammography.

are compared (13,14), and the number of modalities used to identify a cancer (15,16). Population-based density data for more than 300 000 American women have demonstrated that, compared with women of “average” breast density (approximately halfway between scattered areas of fibroglandular density and heterogeneously dense [2], which represents the most clinically relevant approach), the reduction in sensitivity is approximately 7 percentage points for the 40% of women with heterogeneously dense breasts and approximately 13 percentage points for the 10% of women with extremely dense breasts (13). This reduction in mammographic sensitivity is a major contributor to the impetus for supplementary screening modalities.

An additional effect of increased mammographic breast density is an increase in breast cancer risk. The impact of density on breast cancer risk may be misinterpreted as overly important when reviewing studies that describe the risk by comparing the 10% of women with extremely dense breasts to the 10% of women with almost entirely fatty breasts (17–23). This is less meaningful to the overwhelming majority of women because the risk comparison is based on such a small population subset at the two extremes of the density spectrum. When risk is expressed relative to average breast density, the risk for the 40% of women with heterogeneously dense breasts is about 1.2 times greater than average and the risk for the 10% of women with extremely dense breasts is about 2.1 times greater (17). Therefore, breast density is a risk factor, but not a strong one. For example, the risk for a woman

with extremely dense breasts is similar to that for a woman with one first-degree relative with unilateral postmenopausal breast cancer. Furthermore, it makes little sense to consider half of the population undergoing screening mammography at high risk for breast cancer. Overall, CBDIG believes the masking effect of breast density is likely of greater import than the increase in breast cancer risk associated with density alone.

CBDIG Position on Supplementary Screening Modalities and a Risk-based Approach

Mammography is the best single modality for population-based screening (24). It is the only modality proved to significantly reduce mortality from breast cancer in large randomized controlled trials (25–27). Those trials included women of all breast densities and were randomized independent of breast density. Thus, any supplementary screening should be obtained in addition to (not instead of) screening mammography, in accordance with nationally recognized guidelines (28–30).

Supplementary screening tests under consideration for widespread use include breast magnetic resonance (MR) imaging, screening breast ultrasonography (US), and tomosynthesis. In the general population, both US and, to a greater extent, MR imaging provide increased cancer detection over that with mammography alone (31). However, compared with mammography, screening US is associated with a much higher rate of benign biopsies (31–34) and both MR imaging and US result in a much higher rate of recommendation for short-interval follow-up (17,34,35). Although not as widely studied, tomosynthesis is currently being introduced into many radiology practices and preliminary data are encouraging. Population-based screening trials suggest that tomosynthesis may increase breast cancer detection similar to US (albeit not as much as MR imaging) and that tomosynthesis decreases the rate of false-positive findings, even below that seen in screening mammography (36,37).

The radiation dose of the combined two-dimensional plus three-dimensional (3D) mammography examination (as is required for all tomosynthesis examinations) is approximately double that of two-dimensional mammography alone. However, this dose still falls below U.S. Food and Drug Administration (FDA) limits and dose reduction strategies are being actively developed. In particular, the use of synthesized two-dimensional mammographic images created from 3D data has received recent FDA approval, resulting in substantial dose reduction. Thus, the dose-related risk implications for women are considered acceptable.

The California breast density law seeks to promote discussion between women with dense breasts and their physicians regarding the advisability of supplemental screening. In making this decision, the benefit of early cancer detection versus the drawback of increased false-positive findings should be considered. The higher the cancer risk, the more likely there will be benefit from supplemental screening. The benefit-versus-drawback assessment will be more favorable for women who are at high risk on the basis of multiple factors than it will be for average-risk women who only have dense breasts. Thus, CBDIG recommends an individualized risk-based approach for guiding the decision-making process with regard to supplementary screening.

In agreement with the American Cancer Society and National Comprehensive Cancer Network, CBDIG supports the use of validated mathematic models to determine a patient's breast cancer risk (28,30). Proper use of these models requires that the health care provider be fully informed of their merits and weaknesses. Providers can perform a risk assessment or refer the patient to a specialized program. However, given the estimated millions of women who will receive density notification letters, the time demands for detailed risk assessment are likely to be unmanageable. Therefore, it may be valuable for clinicians to elicit "red flag" risk factors to rapidly triage patients. Most important are the known

high-risk factors (eg, *BRCA* genetic mutation). Apart from these, the strongest risk factors for breast cancer are personal or family history (especially at least one first-degree relative with premenopausal breast or ovarian cancer) and personal history of atypia at previous biopsy. Although none of these risk factors, including dense breasts, individually place a woman in the high-risk category, they may identify those who would benefit from a more complete risk assessment.

For the small number of asymptomatic patients who are identified as high risk on the basis of complete risk assessment, supplementary screening is recommended. The American Cancer Society, American College of Radiology, and National Comprehensive Cancer Network recommend screening breast MR imaging annually in addition to yearly mammography for these patients (28,29,38). Although the California legislature did not mandate insurance coverage for any supplementary breast cancer screening tests, screening MR imaging is generally reimbursed for women who are at a lifetime risk of greater than 20%. Some studies have provided support for screening US for high-risk women, but only for those who have no access to or cannot undergo MR imaging (29,31). If a woman undergoes screening MR imaging, screening US will provide no additional benefit (39–43). In addition, many facilities either do not offer screening US or offer it with out-of-pocket charges to the patient.

For patients who are not at high risk, including women with no risk factors other than dense breasts, the pretest probability of breast cancer is low. Therefore, the benefit of supplementary screening is diminished, whereas the potential drawbacks remain the same.

Supplementary screening of the general population of women with heterogeneously dense and extremely dense breasts remains controversial. Among Connecticut women who received a density notification letter, three small studies have shown that screening US helped identify 3.2, 3.2, and 1.8 mammographically occult

cancers for every 1000 women undergoing prevalence (first-time) US screening (35,44,45). Most cancers found with screening US are invasive and at an early stage. This suggests a benefit to screening US. However, there are several major limitations to the currently available data. First, no studies were performed with control groups and no study has provided long-term follow-up. Thus, we do not know the clinical impact of finding these additional small cancers in an average-risk population—specifically whether the cancers would otherwise be detected at the next mammography screening examination while still small, node-negative, and at an early stage and whether there is any associated mortality reduction. A second limitation, as noted previously, is the higher false-positive biopsy rates for lesions identified at US compared with those identified at mammography. Third, no studies in this general patient population have reported outcomes for incidence (subsequent) screening US, although a reduction in the cancer detection rate would be expected (31). This is an important consideration because most breast cancer screening examinations involve incidence rather than prevalence screening. Finally, the outcomes for screening US will likely be more favorable in centers with a dedicated program. Although future large-scale studies will likely address a number of these concerns, they remain pertinent based on currently published literature.

Summary

Breast density notification legislation is becoming increasingly prevalent. The impact of these laws is far reaching, and radiologists should take a proactive role in assessing the need for potential change in their practices and in translating currently available outcomes data into information for referring clinicians and patients that can be easily understood and accessed. In our era of patient-centered care and personalized medicine, breast density notification legislation provides an opportunity for radiologists to engage with referring

clinicians and patients. Statewide collaborations like CBDIG can assist in developing broad-scope guidelines and educational materials, which may minimize the burden for individual breast imaging facilities. The multi-institutional, multidisciplinary CBDIG approach may be a method for organizations to frame responses to individual state laws as similar legislation is passed across the United States.

Disclosures of Conflicts of Interest: E.R.P. No relevant conflicts of interest to disclose. J.H. No relevant conflicts of interest to disclose. J.A.L. No relevant conflicts of interest to disclose. E.A.S. No relevant conflicts of interest to disclose. R.J.B. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: is a consultant for Hologic; is a shareholder in Hologic. Other relationships: none to disclose. K.K.L. No relevant conflicts of interest to disclose. B.N.J. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: was a guest editor for Elsevier MRI Clinics; receives royalties from UpToDate. Other relationships: none to disclose. J.W.T.L. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: receives speaker's honoraria from Hologic; is on the scientific advisory board of Hologic. Other relationships: none to disclose. S.A.F. No relevant conflicts of interest to disclose. L.W.B. No relevant conflicts of interest to disclose. H.O. No relevant conflicts of interest to disclose. B.L.D. Financial activities related to the present article: none to disclose. Financial activities not related to the present article: received grants from GE Healthcare, IFOS, and Intuitive Surgical. Other relationships: none to disclose. A.W.K. No relevant conflicts of interest to disclose. E.L. No relevant conflicts of interest to disclose. L.R. No relevant conflicts of interest to disclose. D.D.W. No relevant conflicts of interest to disclose. D.M.I. No relevant conflicts of interest to disclose.

References

- D'Orsi CJ, Mendelson EB, Ikeda DM, et al. Breast imaging reporting and data system: ACR BI-RADS—breast imaging atlas. Reston, Va: American College of Radiology, 2003.
- D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA. Breast imaging reporting and data system: ACR BI-RADS—breast imaging atlas. 5th ed. Reston, Va: American College of Radiology, 2013 (in press).
- U.S. Census Bureau. State and county quickfacts. <http://quickfacts.census.gov/qfd/states/06000.html>. Accessed March 11, 2013.
- Mammography Quality Standards Act and Programs; US Food and Drug Administration. <http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/default.htm>. Accessed March 11, 2013.
- NCI-funded Breast Cancer Surveillance Consortium. (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040, HHSN261201100031C). Breast Cancer Surveillance Consortium Web site. http://breastscreening.cancer.gov/data/mammography_data.html. Accessed March 11, 2013.
- Sickles EA, Miglioretti DL, Ballard-Barbash R, et al. Performance benchmarks for diagnostic mammography. *Radiology* 2005;235(3):775–790.
- Egan RL, Mosteller RC. Breast cancer mammography patterns. *Cancer* 1977;40(5):2087–2090.
- Whitehead J, Carile T, Kopecky KJ, et al. Wolfe mammographic parenchymal patterns: a study of the masking hypothesis of Egan and Mosteller. *Cancer* 1985;56(6):1280–1286.
- van Gils CH, Otten JD, Verbeek AL, Hendriks JH. Mammographic breast density and risk of breast cancer: masking bias or causality? *Eur J Epidemiol* 1998;14(4):315–320.
- Buist DS, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40–49 years. *J Natl Cancer Inst* 2004;96(19):1432–1440.
- Jackson VP, Hendrick RE, Feig SA, Kopans DB. Imaging of the radiographically dense breast. *Radiology* 1993;188(2):297–301.
- Pinsky RW, Helvie MA. Mammographic breast density: effect on imaging and breast cancer risk. *J Natl Compr Canc Netw* 2010;8(10):1157–1164; quiz 1165.
- Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med* 2003;138(3):168–175.
- Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. *JAMA* 1996;276(1):33–38.
- Corsetti V, Houssami N, Ghirardi M, et al. Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. *Eur J Cancer* 2011;47(7):1021–1026.

16. Leconte I, Feger C, Galant C, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. *AJR Am J Roentgenol* 2003;180(6):1675-1679.
17. Sickles EA. The use of breast imaging to screen women at high risk for cancer. *Radiol Clin North Am* 2010;48(5):859-878.
18. Boyd NF, Byng JW, Jong RA, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *J Natl Cancer Inst* 1995;87(9):670-675.
19. Byrne C, Schairer C, Wolfe J, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *J Natl Cancer Inst* 1995;87(21):1622-1629.
20. Ursin G, Ma H, Wu AH, et al. Mammographic density and breast cancer in three ethnic groups. *Cancer Epidemiol Biomarkers Prev* 2003;12(4):332-338.
21. Vacek PM, Geller BM. A prospective study of breast cancer risk using routine mammographic breast density measurements. *Cancer Epidemiol Biomarkers Prev* 2004;13(5):715-722.
22. Boyd NF, Martin LJ, Sun L, et al. Body size, mammographic density, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2006;15(11):2086-2092.
23. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356(3):227-236.
24. Eddy D. Screening for breast cancer. In: Eddy D, ed. *Common screening tests*. Philadelphia, Pa: American College of Physicians, 1991; 229-254.
25. Tabár L, Vitak B, Chen TH, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology* 2011;260(3):658-663.
26. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AM, Chen TH. The randomized trials of breast cancer screening: what have we learned? *Radiol Clin North Am* 2004;42(5):793-806, v.
27. Smith RA, Brooks D, Cokkinides V, Saslow D, Brawley OW. Cancer screening in the United States, 2013: a review of current American Cancer Society guidelines, current issues in cancer screening, and new guidance on cervical cancer screening and lung cancer screening. *CA Cancer J Clin* 2013;63(2):88-105.
28. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007;57(2):75-89.
29. Mainiero MB, Lourenco A, Mahoney MC, et al. ACR appropriateness criteria breast cancer screening. *J Am Coll Radiol* 2013;10(1):11-14.
30. Bevers TB, Anderson BO, Bonaccio E, et al. NCCN clinical practice guidelines in oncology: breast cancer screening and diagnosis. *J Natl Compr Canc Netw* 2009;7(10):1060-1096.
31. Berg WA, Zhang Z, Lehrer D, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA* 2012;307(13):1394-1404.
32. Nothacker M, Duda V, Hahn M, et al. Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue—a systematic review. *BMC Cancer* 2009;9:335.
33. Sardanelli F, Podo F, Santoro F, et al. Multicenter surveillance of women at high genetic breast cancer risk using mammography, ultrasonography, and contrast-enhanced magnetic resonance imaging (the high breast cancer risk Italian 1 study): final results. *Invest Radiol* 2011;46(2):94-105.
34. Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA* 2008;299(18):2151-2163.
35. Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. *Radiology* 2012;265(1):59-69.
36. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* 2013;267(1):47-56.
37. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 2013;14(7):583-589.
38. Lee CH, Dershaw DD, Kopans D, et al. Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *J Am Coll Radiol* 2010;7(1):18-27.
39. Berg WA. Supplemental screening sonography in dense breasts. *Radiol Clin North Am* 2004;42(5):845-851, vi.
40. Kuhl CK, Schrading S, Leutner CC, et al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *J Clin Oncol* 2005;23(33):8469-8476.
41. Lehman CD, Isaacs C, Schnall MD, et al. Cancer yield of mammography, MR, and US in high-risk women: prospective multi-institution breast cancer screening study. *Radiology* 2007;244(2):381-388.
42. Sardanelli F, Podo F, D'Agno G, et al. Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRI study): interim results. *Radiology* 2007;242(3):698-715.
43. Warner E, Plewes DB, Hill KA, et al. Surveillance of *BRCA1* and *BRCA2* mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA* 2004;292(11):1317-1325.
44. Weigert J, Steenbergen S. The connecticut experiment: the role of ultrasound in the screening of women with dense breasts. *Breast J* 2012;18(6):517-522.
45. Parris T, Wakefield D, Frimmer H. Real world performance of screening breast ultrasound following enactment of Connecticut Bill 458. *Breast J* 2013;19(1):64-70.



4511 North Himes Ave., Suite 250
Tampa, FL 33614

(800) 717-3117
(813) 872-7835
Fax: (813) 873-7838

info@obesityaction.org
www.ObesityAction.org

March 15, 2014

Health Care Authority
626 8th Ave SE
P.O. Box 42712
Olympia, WA 98504-2712

Attention: Christine Masters at shtap@hca.wa.gov

Dear Members of the Washington State Health Care Authority:

On behalf of the nearly 50,000 members of the Obesity Action Coalition (OAC), I am pleased to provide comments in response to the February 28, 2014 notice regarding proposed topics for review by the Washington State Health Care Authority's (HCA) Health Technology Assessment (HTA) program. The OAC's comments will focus on bariatric surgery for those affected by obesity.

Evidence-based literature clearly demonstrate that people affected by obesity can substantially improve their health and quality of life when they have access to a continuum of medically necessary treatment – including behavioral, nutritional, pharmaceutical, psychosocial and surgical treatment. Even a 5 to 10 percent weight-loss produces clinically significant reductions in risk factors for chronic diseases such as diabetes, hypertension, arthritis, heart disease, mental illness, lipid disorders, pulmonary disease (obstructive sleep apnea and restrictive lung disease), and certain cancers.

Throughout its existence, the OAC has learned from our membership that there is no "one-size-fits-all approach to treating obesity." Nationwide, OAC members are advocating on the state and federal level for access to all evidence-based treatment options for the disease of obesity. More than 40 states across the country provide bariatric surgery coverage for their state employees and nearly every state Medicaid program offers this benefit for its low-income residents. We are pleased that Washington State can be counted among these states that recognize the benefits associated with bariatric surgery for their state employees and Medicaid recipients.

Therefore, it is our hope that the Washington State HCA is simply evaluating the growing evidence in support of metabolic and bariatric surgery and the tremendous possibilities that this particular intervention holds for those in the early stages of obesity (body mass index between 30 and 34.9 kg/m²). The OAC encourages HCA members to carefully review the evidence and practice guidelines that are included in the American Society for Metabolic and Bariatric Surgery's formal comments on the February 28th notice. We believe that this evidence clearly speaks to the necessity of maintaining the bariatric surgery benefit for Washington state employees.

On behalf of the more than 900 OAC members in the State of Washington, I thank you for your consideration of the OAC's comments on this issue. Should you have any questions or need additional information, please feel free to contact either me, or the OAC's Washington Policy Consultant Christopher Gallagher at 571-235-6475 or via email at chris@potomaccurrents.com. Thank you.

Sincerely,

A handwritten signature in black ink, which appears to be "Joe Nadglowski". Below the signature is a solid black rectangular redaction box.

Joe Nadglowski
President & CEO
Obesity Action Coalition

Masters, Christine V. (HCA)

From: Peery, Harrison <Hpeery@entnet.org>
Sent: Monday, March 17, 2014 12:31 PM
To: HCA ST Health Tech Assessment Prog
Cc: Kappel, Jenna; Minton, Jenna; Jarchow, Danielle; Brereton, Jean; Masters, Christine V. (HCA)
Subject: Re: WHA Topics for Review in 2015: Imaging for Rhinosinusitis and Tympanostomy Tubes
Importance: High

Dear Ms. Masters:

The American Academy of Otolaryngology – Head and Neck Surgery (AAO-HNS) represents approximately 12,000 physicians in the United States who diagnose and treat disorders of the ears, nose, throat, and related structures of the head and neck. The medical ailments treated by this specialty are the most common that afflict all Americans, old and young, including hearing loss, balance disorders, chronic ear infections, rhinological disorders, snoring and sleep disorders, swallowing disorders, facial and cranial nerve disorders, and head and neck cancer.

After reviewing the Washington Healthcare Authority's (WHA's) most recent selections for health technologies to undergo review in 2015, we have identified several AAO-HNS resources that are directly relevant to WHA's review of two topics: 1) Imaging for Rhinosinusitis and 2) Tympanostomy Tubes. We hope that WHA will consider the following AAO-HNS resources as evidence for consideration during review of these two important topics:

Imaging for Rhinosinusitis:

- Clinical Practice Guideline on Adult Sinusitis (Update in Progress): http://www.entnet.org/guide_lines/Adult-Sinusitis.cfm
- Clinical Consensus Statement on Imaging for Paranasal Sinus Disease: <http://oto.sagepub.com/content/147/5/808.abstract>
- Clinical Indicator Adult Sinus Surgery: <http://www.entnet.org/Practice/Endoscopic-Sinus-Surgery-Adult.cfm>
- Clinical Indicator Pediatric Sinus Surgery: <http://www.entnet.org/Practice/Endoscopic-Sinus-Surgery-Pediatric.cfm>
- Clinical Indicator Diagnostic Nasal Endoscopy: <http://www.entnet.org/Practice/Diagnostic-Nasal-Endoscopy.cfm>
- Position Statement on Intra-Operative Use of Computer Guided Imagery <http://www.entnet.org/Practice/policyIntraOperativeSurgery.cfm>
- Position Statement on Sinus Endoscopy: <http://www.entnet.org/Practice/policySinusEndoscopy.cfm>
- Position Statement on Point of Care Imaging: <http://www.entnet.org/Practice/policyReimburseImagingStudies.cfm>
- Patient Fact Sheet on Allergic Rhinitis, Sinusitis and Rhinosinusitis: <http://www.entnet.org/HealthInformation/rhinitis.cfm>

Tympanostomy Tubes:

- Clinical Practice Guideline: Tympanostomy Tubes in Children: http://www.entnet.org/guide_lines/Tubes.cfm
- Clinical Indicator on Myringotomy and Tympanostomy Tubes: <http://www.entnet.org/Practice/Myringotomy-and-Tympanostomy-Tubes.cfm>
- Position Statement on Middle-Ear Ventilation Tube Placement: <http://www.entnet.org/Practice/policyMidEarVentPlacement.cfm>

If you have further information regarding the schedule and process of these two reviews, please feel free to forward on and keep us posted as we would like to submit additional comments and potentially have one of our physician leaders attend the meetings as we have done in the past. Please let me know if you have any questions about any of the above listed resources.

Best Regards,

Harrison Peery
 Health Policy Analyst
 American Academy of Otolaryngology - Head and Neck Surgery
 P: 703-535-3728
 E: hpeery@entnet.org

From: Masters, Christine V. (HCA) [<mailto:christine.masters@hca.wa.gov>]
Sent: Friday, February 28, 2014 8:31 PM
Subject: HTA Program Update: Six topics chosen for future review

The Health Care Authority Director has selected five health technologies to undergo initial review and coverage decisions by the Health Technology Clinical Committee (HTCC) beginning in 2015. One additional topic has been selected for re-review by the HTCC. The list of selected topics appears below and on the [HTA website](#).

Topics Selected for Review:

1. Testosterone Testing
2. Bariatric Surgery for Overweight/ Obese
3. Imaging for Rhinosinusitis
4. Appropriate Breast Imaging for Breast Cancer Screening in Special Conditions
5. Tympanostomy Tubes

Topics Selected for Re-review:

1. Lumbar Fusion

To gather information and evidence for consideration during review of the selected topics, public comments will be accepted from **February 28th until 5 p.m. March 17, 2014**. Please submit all comments to: shtap@hca.wa.gov.

For more information about the Health Technology Assessment (HCA) Program or the topic selection process, visit our [website](#).

Christine Masters | Program Specialist
 Health Technology Assessment Program
 626 8th Avenue SE | Olympia, WA 98501
 Desk 360-725-5126 | christine.masters@hca.wa.gov



Michael Bolen
Director, State Government
Affairs
Medtronic, Inc.
1424 Royal Oak Dr.
Blue Bell, PA 19422

tel 215 499 1379
michael.bolen@medtronic.co

**Response to the Washington State Health Care Authority's Health Technology Assessment
for 2001101 – Lumbar Fusion**

March 17, 2014

Washington State Healthcare Authority
Health Technology Assessment Program
P.O. Box 42712
Olympia, WA 98504-2712

Re: Potential 2014 HTA Technology Topics

To Whom It May Concern:

We appreciate the opportunity to comment on the Potential 2014 HTA Technology Topics and in particular the inclusion for re-review of the guideline for *Lumbar Fusion for Uncomplicated Degenerative Disc Disease*. We understand the challenges of maintaining clinical guidelines and we applaud your efforts to methodologically sound process.

We have conducted a PubMed search, using the “key words” identified in your announcements, and have found numerous published studies since 2012. A bibliography of these articles is enclosed for your reference. In particular, we would like to direct your attention to a meta-analysis by Phillips, et al., which found that lumbar fusion is a viable treatment to patients that fail conservative management.

Additionally, we want to make you aware of a recent review of the same topic by the Agency for Healthcare Research and Quality (AHRQ) which included six randomized control trials (RCTs), which found, overall, positive results when lumbar fusion was performed in this patient population. In addition, we believe that any proper guideline should not rely solely on the published evidence, but should also consider expert opinion from the appropriate specialty societies. Both the American Association of Neurological Surgeons (AANS) and the International Society for the Advancement of Spine Surgery (ISASS) have expressed opinions on this topic in response to the review conducted by AHRQ. We have enclosed a copy of the published letter from AANS and we have provided a web link to the policy statement on lumbar spinal fusion surgery by ISASS.

While we have not addressed all the published data in our comments, we certainly look forward to your report on the potential of revised guideline and the opportunity to comment and/or discussion in person before the finalization of any revisions. Thank you again for the opportunity to comment.

Sincerely,

(Sent Electronically)

Michael J. Bolen

Director, Government Affairs



SEARCH QUERY
(RCTs) [PD March 14



Phillips Fusion
MetaAnalysis.pdf



AANS Comments to
AHRQ - Spinal Fusion

http://www.isass.org/public_policy/2011-07-15_policy_statement_lumbar_surgery.html

SEARCH QUERY

"lumbar"[tw] AND ("DDD"[tw] OR "degenerative disc"[tw]) AND ("fusion"[tw] OR "arthrodesis"[tw]) AND (outcome* OR "effectiveness"[tw] OR "efficacy"[tw]) AND (Clinical Trial[ptyp] OR Comparative Study[ptyp] OR Controlled Clinical Trial[ptyp] OR Meta-Analysis[ptyp] OR Randomized Controlled Trial[ptyp] OR Review[ptyp] OR systematic[sb]) AND ("2012/09/01"[PDat] : "2014/03/13"[PDat]) AND Humans[Mesh] AND English[lang]

SEARCH RESULTS

1. Choi UY, Park JY, Kim KH, Kuh SU, Chin DK, Kim KS, Cho YE. Unilateral versus bilateral percutaneous pedicle screw fixation in minimally invasive transforaminal lumbar interbody fusion. *Neurosurg Focus*. 2013 Aug;35(2):E11. doi: 10.3171/2013.2.FOCUS12398. PubMed PMID: 23905949.

2. Lau D, Khan A, Terman SW, Yee T, La Marca F, Park P. Comparison of perioperative outcomes following open versus minimally invasive transforaminal lumbar interbody fusion in obese patients. *Neurosurg Focus*. 2013 Aug;35(2):E10. doi: 10.3171/2013.5.FOCUS13154. PubMed PMID: 23905948.

3. Wei J, Song Y, Sun L, Lv C. Comparison of artificial total disc replacement versus fusion for lumbar degenerative disc disease: a meta-analysis of randomized controlled trials. *Int Orthop*. 2013 Jul;37(7):1315-25. doi: 10.1007/s00264-013-1883-8. Epub 2013 May 4. PubMed PMID: 23645003; PubMed Central PMCID: PMC3685666.

4. Papavero L, Kothe R. Microsurgical extraforaminal decompression of lumbar root canal stenosis. *Oper Orthop Traumatol*. 2013 Feb;25(1):16-30. doi: 10.1007/s00064-012-0194-3. PubMed PMID: 23381738.

5. Phillips FM, Slosar PJ, Youssef JA, Andersson G, Papatheofanis F. Lumbar spine fusion for chronic low back pain due to degenerative disc disease: a systematic review. *Spine (Phila Pa 1976)*. 2013 Apr 1;38(7):E409-22. doi: 10.1097/BRS.0b013e3182877f11. Review. PubMed PMID: 23334400.

6. Dahl MC, Ellingson AM, Mehta HP, Huelman JH, Nuckley DJ. The biomechanics of a multilevel lumbar spine hybrid using nucleus replacement in conjunction with fusion. *Spine J*. 2013 Feb;13(2):175-83. doi: 10.1016/j.spinee.2012.11.045. Epub 2013 Jan 11. PubMed PMID: 23318109.

7. Zairi F, Arikat A, Allaoui M, Assaker R. Transforaminal lumbar interbody fusion: comparison between open and mini-open approaches with two years follow-up. *J Neurol Surg A Cent Eur Neurosurg*. 2013 May;74(3):131-5. doi: 10.1055/s-0032-1330956. Epub 2013 Jan 10. PubMed PMID: 23307308.

8. Patel VC, Park DK, Herkowitz HN. Lateral transpsoas fusion: indications and outcomes. *ScientificWorldJournal*. 2012;2012:893608. doi: 10.1100/2012/893608. Epub 2012 Nov 6. Review. PubMed PMID: 23213303; PubMed Central PMCID: PMC3504425.

9. Burkus JK, Dryer RF, Pelozo JH. Retrograde ejaculation following single-level anterior lumbar surgery with or without recombinant human bone morphogenetic protein-2 in 5 randomized controlled trials: clinical article. *J Neurosurg Spine*. 2013 Feb;18(2):112-21. doi: 10.3171/2012.10.SPINE11908. Epub 2012 Nov 30. PubMed PMID: 23199378.
10. Zigler JE, Glenn J, Delamarter RB. Five-year adjacent-level degenerative changes in patients with single-level disease treated using lumbar total disc replacement with ProDisc-L versus circumferential fusion. *J Neurosurg Spine*. 2012 Dec;17(6):504-11. doi: 10.3171/2012.9.SPINE11717. Epub 2012 Oct 19. PubMed PMID: 23082849.
11. Zigler JE, Delamarter RB. Five-year results of the prospective, randomized, multicenter, Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential arthrodesis for the treatment of single-level degenerative disc disease. *J Neurosurg Spine*. 2012 Dec;17(6):493-501. doi: 10.3171/2012.9.SPINE11498. Epub 2012 Oct 19. PubMed PMID: 23082846.
12. Jacobs WC, van der Gaag NA, Kruyt MC, Tuschel A, de Kleuver M, Peul WC, Verbout AJ, Oner FC. Total disc replacement for chronic discogenic low back pain: a Cochrane review. *Spine (Phila Pa 1976)*. 2013 Jan 1;38(1):24-36. doi: 10.1097/BRS.0b013e3182741b21. Review. PubMed PMID: 22996268.
13. Jacobs W, Van der Gaag NA, Tuschel A, de Kleuver M, Peul W, Verbout AJ, Oner FC. Total disc replacement for chronic back pain in the presence of disc degeneration. *Cochrane Database Syst Rev*. 2012 Sep 12;9:CD008326. doi: 10.1002/14651858.CD008326.pub2. Review. PubMed PMID: 22972118.
14. Lawrence BD, Wang J, Arnold PM, Hermsmeyer J, Norvell DC, Brodke DS. Predicting the risk of adjacent segment pathology after lumbar fusion: a systematic review. *Spine (Phila Pa 1976)*. 2012 Oct 15;37(22 Suppl):S123-32. doi: 10.1097/BRS.0b013e31826d60d8. PubMed PMID: 22885827.
15. Duncan JW, Bailey RA. An analysis of fusion cage migration in unilateral and bilateral fixation with transforaminal lumbar interbody fusion. *Eur Spine J*. 2013 Feb;22(2):439-45. doi: 10.1007/s00586-012-2458-x. Epub 2012 Aug 10. PubMed PMID: 22878377; PubMed Central PMCID: PMC3555613.
16. Wang JC, Arnold PM, Hermsmeyer JT, Norvell DC. Do lumbar motion preserving devices reduce the risk of adjacent segment pathology compared with fusion surgery? A systematic review. *Spine (Phila Pa 1976)*. 2012 Oct 15;37(22 Suppl):S133-43. doi: 10.1097/BRS.0b013e31826cadf2. Review. PubMed PMID: 22872221.
17. Daftari TK, Chinthakunta SR, Ingalhalikar A, Gudipally M, Hussain M, Khalil S. Kinematics of a selectively constrained radiolucent anterior lumbar disc: comparisons to hybrid and circumferential fusion. *Clin Biomech (Bristol, Avon)*. 2012 Oct;27(8):759-65. doi: 10.1016/j.clinbiomech.2012.05.010. Epub 2012 Jun 27. PubMed PMID: 22748372.

18. Siepe CJ, Heider F, Haas E, Hitzl W, Szeimies U, Stäbler A, Weiler C, Nerlich AG, Mayer MH. Influence of lumbar intervertebral disc degeneration on the outcome of total lumbar disc replacement: a prospective clinical, histological, X-ray and MRI investigation. *Eur Spine J*. 2012 Nov;21(11):2287-99. doi: 10.1007/s00586-012-2342-8. Epub 2012 May 29. PubMed PMID: 22644434; PubMed Central PMCID: PMC3481086.

19. Oestergaard LG, Nielsen CV, Bünger CE, Sogaard R, Fruensgaard S, Helmig P, Christensen FB. The effect of early initiation of rehabilitation after lumbar spinal fusion: a randomized clinical study. *Spine (Phila Pa 1976)*. 2012 Oct 1;37(21):1803-9. PubMed PMID: 22565381.

20. Pelton MA, Phillips FM, Singh K. A comparison of perioperative costs and outcomes in patients with and without workers' compensation claims treated with minimally invasive or open transforaminal lumbar interbody fusion. *Spine (Phila Pa 1976)*. 2012 Oct 15;37(22):1914-9. doi: 10.1097/BRS.0b013e318257d490. PubMed PMID: 22487713.

21. Van de Kelft E, Verguts L. Clinical outcome of monosegmental total disc replacement for lumbar disc disease with ball-and-socket prosthesis (Maverick): prospective study with four-year follow-up. *World Neurosurg*. 2012 Sep-Oct;78(3-4):355-63. doi: 10.1016/j.wneu.2011.10.043. Epub 2011 Dec 11. PubMed PMID: 22120556.

22. Strube P, Hoff E, Hartwig T, Perka CF, Gross C, Putzier M. Stand-alone anterior versus anteroposterior lumbar interbody single-level fusion after a mean follow-up of 41 months. *J Spinal Disord Tech*. 2012 Oct;25(7):362-9. PubMed PMID: 21716142.

ATTACHMENTS

LITERATURE REVIEW

Lumbar Spine Fusion for Chronic Low Back Pain Due to Degenerative Disc Disease

A Systematic Review

Frank M. Phillips, MD,* Paul J. Slosar, MD,† Jim A. Youssef, MD,‡ Gunnar Andersson, MD, PhD,* and Frank Papatheofanis, MD, PhD§

Study Design. Systematic literature review.

Objective. To categorize published evidence systematically for lumbar fusion for chronic low back pain (LBP) in order to provide an updated and comprehensive analysis of the clinical outcomes.

Summary of Background Data. Despite a large number of publications of outcomes of spinal fusion surgery for chronic LBP, there is little consensus on efficacy.

Methods. A MEDLINE and Cochrane database search was performed to identify published articles reporting on validated patient-reported clinical outcomes measures (2 or more of visual analogue scale, Oswestry Disability Index, Short Form [36] Health Survey [SF-36] PCS, and patient satisfaction) with minimum 12 months of follow-up after lumbar fusion surgery in adult patients with LBP due to degenerative disc disease. Twenty-six total articles were identified and stratified by level of evidence: 18 level 1 (6 studies of surgery vs. nonoperative treatment, 12 studies of alternative surgical procedures), 2 level 2, 2 level 3, and 4 level 4 (2 prospective, 2 retrospective). Weighted averages of each outcomes measure were computed and compared with established minimal clinically important difference values.

Results. Fusion cohorts included a total of 3060 patients. The weighted average improvement in visual analogue scale back pain was 36.8/100 (standard deviation [SD], 14.8); in Oswestry Disability Index 22.2 (SD, 14.1); in SF-36 Physical Component Scale 12.5 (SD, 4.3). Patient satisfaction averaged 71.1% (SD, 5.2%) across studies. Radiographical fusion rates averaged 89.1% (SD, 13.5%),

and reoperation rates 12.5% (SD, 12.4%) overall, 9.2% (SD, 7.5%) at the index level. The results of the collective studies did not differ statistically in any of the outcome measures based on level of evidence (analysis of variance, $P > 0.05$).

Conclusion. The body of literature supports fusion surgery as a viable treatment option for reducing pain and improving function in patients with chronic LBP refractory to nonsurgical care when a diagnosis of disc degeneration can be made.

Key words: low back pain, disc degeneration, lumbar fusion, outcome. **Spine 2013;38:E409–E422**

Chronic low back pain (LBP), defined as pain lasting more than 3 months, is the second most common reason for visits to a physician and the most common reason for missing work across all socioeconomic strata in the United States.^{1–3} Chronic LBP occurs in 5% to 8% of community-dwelling persons^{4,5} and is reported in 19% of working adults.⁶ The total costs of the condition are estimated at more than \$100 billion annually, with two-thirds of that due to decreased wages and productivity.⁷

Chronic LBP is known to be associated with degeneration of the spinal motion segment.^{1,2,8–10} Degeneration is thought to initiate in the intervertebral disc with subsequent degeneration occurring in the facet joints.⁸ Although disc degeneration occurs frequently with aging, and may be asymptomatic in most, in certain instances it can cause severe LBP.^{1–10}

Activity modification, medications, and physical therapy remain the mainstay of nonoperative management of chronic LBP. Although potentially helpful in the acute setting, the results of physical therapy for chronic LBP are unclear, and there is limited evidence that physical therapy treatments help to prevent recurrent or chronic back pain.^{11,12} Narcotic medications are frequently prescribed and substance use disorders are common in patients with chronic back pain, and aberrant medication-taking behaviors occur in up to 24% of cases.¹³

Patients with debilitating back pain who fail to find relief with nonoperative management, and whose origin of pain is thought to reside within the motion segment may be considered for surgical fusion^{14,15} with the intent of eliminating painful motion. The efficacy of fusion in the treatment of chronic LBP without radiculopathy has been challenged. With reports of increases in spinal surgical costs and volumes,¹⁶ pressures for

From the *Department of Orthopaedic Surgery, Rush University Medical Center, Chicago, IL; †SpineCare Medical Group, San Francisco Spine Institute, San Francisco, CA; ‡Spine Colorado/Durango Orthopedic Associates, P.C., Durango, CO; and §Saint Katherine College, Encinitas, CA.

Acknowledgment date: April 2, 2012. First revision date: July 23, 2012. Second revision date: November 8, 2012. Acceptance date: November 16, 2012.

The manuscript submitted does not contain information about medical device(s)/drug(s).

Research support received from Nuvasive Inc., San Diego, CA.

Relevant financial activities outside the submitted work: consulting fee or honorarium, consultancy, board membership, patents, royalties, stock/stock options, travel/accommodations/meeting expenses.

Address correspondence and reprint requests to Frank M. Phillips, MD, Midwest Orthopaedics at Rush, 1611 W. Harrison St, Ste 300, Chicago, IL 60612; E-mail: frank.phillips@rushortho.com

DOI: 10.1097/BRS.0b013e3182877f11

Spine 2014 Technology Topic Selection

Copyright © 2013 Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

www.spinejournal.com | 5710677 E409

| TABLE 1. Search Criteria | |
|--|--|
| Limits | English (Language) "Humans" (Medical Subject Headings) |
| Search Terms (Inclusion) | |
| Lumbar fusion surgery for low back pain due to degenerative disc disease | fusion [Title/Abstract] OR |
| | surgery [Title/Abstract] OR |
| | surgical [Title/Abstract] AND |
| | lumbar AND |
| | "degenerative disc disease" [Title/Abstract] OR |
| | "degenerative disk disease" [Title/Abstract] OR |
| | "disc degeneration" [Title/Abstract] OR |
| | DDD [Title/Abstract] OR |
| | spondylosis [Title/Abstract] OR |
| "back pain" [Title/Abstract] | |
| Search Terms (Exclusion) | |
| Nonclinical/preclinical/diagnostic reports | NOT "case report*" [Title] |
| | NOT "learning curve" [Title] |
| | NOT biomechan* [Title/Abstract] OR |
| | kinemat* [Title/Abstract] OR |
| | "finite element" [Title/Abstract] |
| | NOT histo* [Title] OR |
| | patho* [Title] OR |
| | psycho* [Title] |
| | NOT cadaver* [Title] OR |
| | anatom* [Title] |
| | NOT geometry [Title] |
| | NOT imaging [Title] OR |
| | MRI [Title] |
| | NOT assessment [Title] OR |
| | diagn* [Title] |
| | Nonfusion surgical reports |
| microdiscectomy [Title] OR | |
| laminectomy [Title] OR | |
| decompress* [Title] | |
| NOT dynamic [Title] | |
| NOT nucle* [Title] | |
| NOT laminoplasty [Title] | |
| NOT kyphoplasty [Title] OR | |

(Continued)

| TABLE 1. (Continued) | |
|--------------------------|--|
| Limits | English (Language) "Humans" (Medical Subject Headings) |
| | vertebroplasty [Title] |
| | NOT osteotomy [Title] |
| Other treatments | NOT injection* [Title] |
| | NOT steroid* [Title] |
| | NOT epidural [Title] OR |
| | dura* [Title] |
| | NOT electro* [Title] OR |
| | electrical [Title] |
| | NOT radio* [Title] |
| Other diagnoses | NOT cervical [Title] OR |
| | thoracic [Title] OR |
| | pelvic [Title] |
| | NOT scolio* [Title] OR |
| | deformity [Title] OR |
| | kyph* [Title] |
| | NOT spondylolisthesis [Title] |
| | NOT lytic [Title] |
| | NOT isthmic [Title] |
| | NOT tumor [Title] OR |
| | carcinoma [Title] |
| | NOT spondylolysis [Title] |
| | NOT traum* [Title] OR |
| | fracture [Title] |
| | NOT infect* [Title] OR |
| | osteomyelitis [Title] OR |
| | spondylitis [Title] OR |
| | tuberculosis [Title] OR |
| | cyst [Title] |
| | NOT stenosis [Title] OR |
| | radicul* [Title] OR |
| | nerve [Title] OR |
| | "cauda equina syndrome" [Title] OR |
| "spinal cord" [Title] OR | |
| sciatica [Title] | |
| NOT adjacent [Title] | |
| NOT children [Title] OR | |
| adolescent [Title] | |

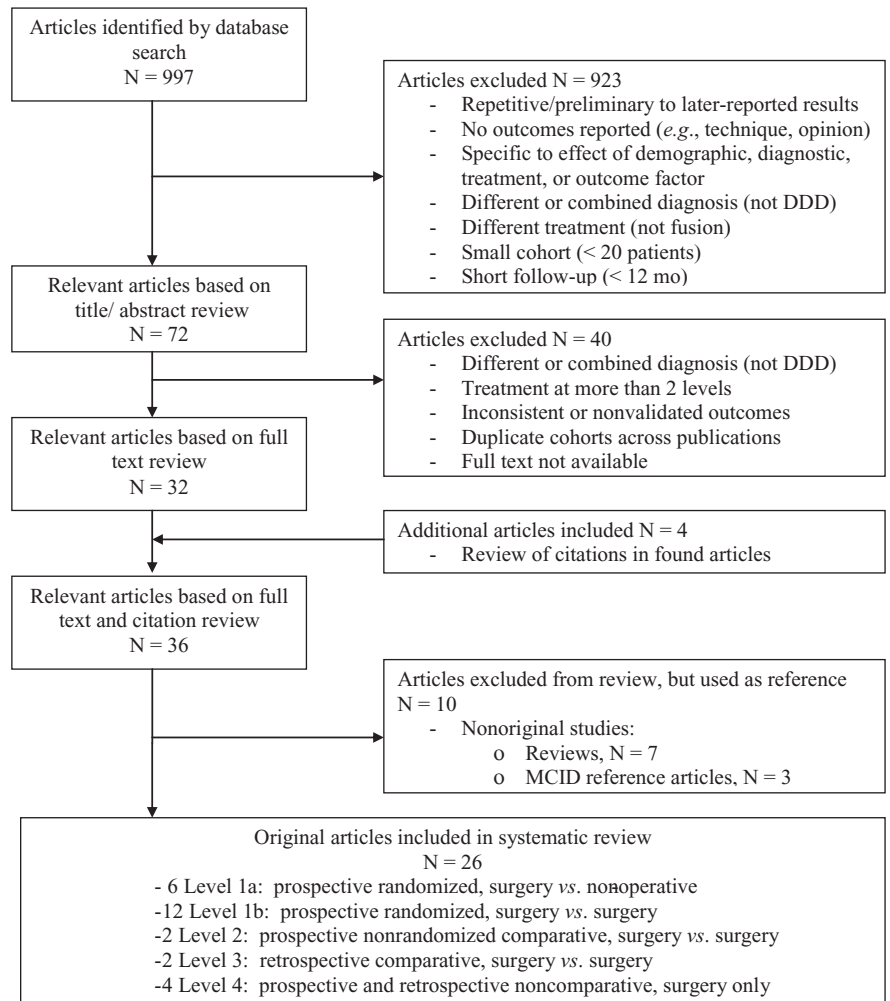


Figure 1. Flow diagram of systematic review. DDD indicates degenerative disc disease; MCID, minimal clinically important difference.

cost containment have led to insurance company policies that limit coverage of fusion procedures. Although undisputedly there is room for more efficient provision of health care spending, the concern is growing that insurers' denials regarding spine fusion surgery are increasingly being based on nontransparent guidelines provided by consulting firms, without input from the physicians, their respective societies, or evidence-based literature.

Most published systematic reviews on the topic of fusion for chronic LBP have focused on randomized controlled trials (RCTs). Although RCTs represent the highest level of evidence available, they are not always feasible or practical to address surgical treatments.^{17,18} The current systematic review was performed to analyze and compile the currently available published literature on fusion for chronic back pain with underlying disc degeneration, updating the evidence with recent studies, and broadening the scope of prior reviews to include a range of study designs beyond RCTs, weighted for evaluation.

MATERIALS AND METHODS

Literature Search

Prospective and retrospective randomized and nonrandomized studies of adults with a primary complaint of LBP

and/or diagnoses of degenerative disc disease (DDD) or spondylosis who were treated by lumbar fusion surgery at 1 or 2 disc levels through any surgical approach were reviewed. Some studies included patients with up to grade 1 spondylolisthesis; however, these were only included if the study reported symptoms of back pain greater than leg pain. A search of the MEDLINE and Cochrane databases from 1950 through July of 2011 was conducted using the search terms outlined in Table 1.

Articles were included only if they reported on a minimum of 20 patients with a minimum of 12 months of follow-up, and included at least 2 of the following validated outcome measures: pain (*via* a numerical rating or visual analogue scale [VAS]), back-specific functional status (Oswestry Disability Index, or ODI), a global measure of generic health improvement (SF-36), or patient satisfaction. Fusion results were also compiled, although methods of determination of fusion varied across studies. Studies were reviewed for overlap of patients and were excluded if they included the same cohort of patients. Complication rates could not be systematically summarized because of inconsistencies in reporting. Reoperation rate was evaluated as a somewhat more reliable measure of negative health outcomes.

TABLE 2. Summary of Prospective Randomized Controlled Trials of Surgery Versus Nonsurgical Treatment (Level 1a)

| Article | Cohorts | n* | Lvls | Follow-up (mo) | Mean Age (yr) | Disco-gram Required | Baseline Back Pain (/100 \ddagger) | Change in Back Pain (abs; [%]) | Baseline ODI (/100) | Change in ODI (abs; [%]) | Baseline SF-36 PCS (/100) | Change in SF-36 PCS (abs; [%]) | Satisf/Redo Rate (%) | Fusion Rate (%) | Reop Rate (%) |
|------------------------------|--|-----|---------|----------------|---------------|------------------------|---------------------------------------|--------------------------------|---------------------|--------------------------|---------------------------|--------------------------------|----------------------|-----------------|---------------|
| Fritzell et al ²³ | PLF, Inst. PLF, ALIF/PLIF | 222 | | | 43 | | 64.2 | 21.0 (33) | 47.3 | 11.6 (25) | NR | NR | 75 | 83 | 7.6 |
| | Common nonsurgical treatments, mainly PT | 72 | 1-2 | 24 | 44 | N | 62.6 | 4.3 (7) | 48.4 | 2.8 (6) | NR | NR | 53 | NA | NA |
| Brox et al ¹⁹ | Inst. PLF | 37 | | | 44.1 | | 62.1 | 22.7 (37) | 42.0 | 15.6 (37) | NR | NR | NR | 84 | NR |
| | Cognitive intervention & exercise | 27 | 1-2 | 12 | 42.4 | N | 64.1 | 15.4 (24) | 43.0 | 13.3 (31) | NR | NR | NR | NA | NA |
| Brox et al ²⁰ | Inst. PLF | 29 | | | 42 | | 64.6 | 13.9 (22) | 47.0 | 8.9 (19) | NR | NR | NR | 85 | NR |
| | Cognitive intervention & exercise | 31 | 1-2 | 12 | 43 | N | 64.7 | 15.2 (23) | 45.1 | 12.8 (28) | NR | NR | NR | NA | NA |
| Fairbank et al ²² | Lumbar fusion | 176 | | | NR | | NR | NR | 46.5 | 12.5 (27) | 19.4 | 9.4 (48) | NR | NR | 6 |
| | Intensive rehab | 173 | 1-2 | 24 | NR | N | NR | NR | 44.8 | 8.7 (19) | 20 | 7.6 (38) | NR | NA | NA |
| Brox et al ²¹ | Inst. PLF | 62 | | | 42.7 | | 63.3 | 22.8 (36) | 44.4 | 15.3 (34) | NR | NR | NR | NR | NR |
| | Cognitive intervention and exercise | 49 | 1-2 | 48 | 43.9 | N | 64.2 | 17.4 (27) | 43.0 | 2.5 (6) | NR | NR | NR | NA | NA |
| Ohtori et al ²⁴ | ALIF (15) and PLF (6) \ddagger | 21 | 1 | 24 | 35.6 | Y | 71.4 | 55.0 (77) | 63.1 | 49.7 (79) | NR | NR | 72 | 100 | NR |
| | Exercise | 20 | 2 | | 33 | | 7.7 | 3.0 (39) | 64 | 24 (38) | NR | NR | 65 | NA | NR |
| Surgery | Total | 547 | Average | | 41.5 | Weighted averages \S | 64.3 | 22.8 (35.3) | 46.9 | 13.9 (29.0) | 19.4 | 9.4 (48.5) | 74.8 | 84.5 | 7.0 |
| | | | SD | | 3.4 | W-SDS | | 10.6 (14.1) | | | | | | | 6.3 |
| Nonop | Total | 372 | Average | | 41.3 | Weighted averages \S | 58.0 | 10.6 (20.0) | 46.2 | 8.2 (17.5) | 20.0 | 7.6 (38.0) | 55.6 | NA | NA |
| | | | SD | | 4.7 | W-SDS | | 7.1 (12.3) | | | | | | | 2.9 |

*The number of patients is reflective of cohorts reported at last follow-up, not enrollment.

\ddagger Fusion cohorts in the study by Ohtori et al²⁴ were combined.

$\#$ VAS scores were converted to a 100-point score where necessary, for consistent interpretation.

\S Where appropriate, outcomes measures were averaged weighted by individual study size. Weighted standard deviations (W-SD) across groups of studies are also presented for change in outcome measures over time.

FU indicates follow-up; ODI, Oswestry Disability Index; PCS, physical component score; PLF, posterolateral fusion; ALIF, anterior lumbar interbody fusion; PLIF, posterior lumbar interbody fusion; PT, physical therapy; NR, not reported; NA, not applicable; SD, standard deviation; W-SD, weighted standard deviation (weighted by study size); SF-36, Short Form (36) Health Survey; n, number of patients; lvls, levels; mo, months; yr, years; abs, absolute; Min, minimum; Max, maximum; abs, absolute value; Satisf, satisfaction; Reop, reoperation (includes nonindex level procedures); Inst, instrumented; rehab, rehabilitation; Nonop, nonoperative.

The search, diagrammed in Figure 1, resulted in 26 original studies, which were then categorized by study design and stratified on the basis of level of evidence. Weighted averages of each outcomes measure were computed for the stratified subgroups of studies; weighting was based on individual study sample size (at last follow-up). Standard deviations and 95% confidence intervals were also computed for the weighted averages. The overall weighted averages across all studies were compared with established minimal clinically important difference (MCID) values.

RESULTS

Studies of Surgical Versus Nonsurgical Treatment

Six publications¹⁹⁻²⁴ (Table 2) report on the results of prospective randomized studies of fusion surgery *versus* nonsurgical therapy in patients with moderately severe pain and disability lasting for at least 1 year who were unresponsive to standard nonsurgical therapy. Inclusion was generally based on presentation of symptomatic chronic LBP and identification of level-specific degenerative changes on radiographical imaging. Patients were excluded if they had signs of neural compression, generalized disc degeneration shown on radiographs, or other specific radiological findings such as stenosis, spondylolisthesis, fracture, infection, or neoplasm. Discography was intermittently used, although as an inclusion criterion in only 1 study.²⁴ Fusion techniques involved most commonly instrumented posterolateral fusion (PLF), but also noninstrumented PLF and lumbar interbody fusion *via* anterior (anterior lumbar interbody fusion [ALIF]) or posterior (posterior lumbar interbody fusion [PLIF]) approach,²³ and ALIF in addition to PLF.²⁴ Patients in the nonsurgical groups continued standard nonoperative care (mainly physical therapy)²³ or received structured rehabilitation including exercise programs and/or cognitive interventions.^{19-22,24}

As a whole, these 6 studies including 547 fusion and 372 nonsurgical patients showed a weighted average improvement in back pain of $22.8 \pm 10.6/100$ points (35.3% change; 95% CI, 23.0–47.6) in the surgical group, compared with $10.6 \pm 7.1/100$ points (20% change; 95% CI, 9.2–30.8) in the nonsurgical group. Weighted average improvement in ODI was $13.9 \pm 8.7/100$ (29% change; 95% CI, 18.7–39.4) in the surgical group and $8.2 \pm 6.2/100$ (17.5% change; 95% CI, 8.5–26.6) in the nonsurgical group (Figure 2). Only 1 study²² reported SF-36 Physical Component Scores (PCS), with a 9.4/100 point improvement (48% change) in the surgical group and 7.6/100 point improvement (38% change) in the nonsurgical group. Weighted average satisfaction rates were 74.8% (95% CI, 72.2–77.4) in the surgical group and 55.6% (95% CI, 53.3–57.9) in the nonsurgical group. Fusion rate in the surgical group was 84% (95% CI, 78.3–90.6). The average reoperation rate in the surgical group was 7% (95% CI, 5.7–8.3).

Studies of Surgery-Only Cohorts: Prospective Randomized Trials

Twelve publications²⁵⁻³⁶ (Table 3) report on the results of prospective randomized studies of fusion surgery comparing

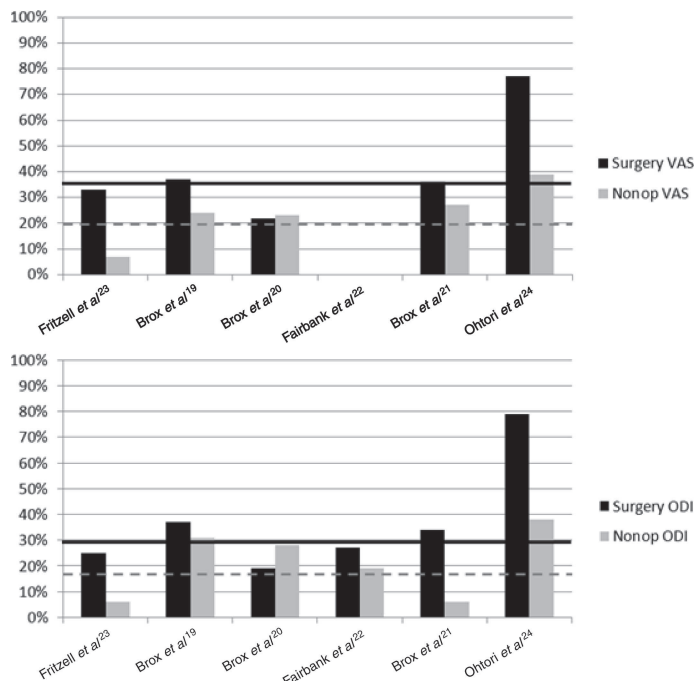


Figure 2. Graphical representation of reported percentage change (from preoperative to last follow-up) in back pain visual analogue scale (VAS) and Oswestry Disability Index (ODI) among studies of surgery *versus* nonoperative treatment.¹⁹⁻²⁴ Solid horizontal lines represent weighted averages across combined cited studies; solid black, surgery; dashed gray, nonoperative. Nonop indicates nonoperative.

technique (*e.g.*, anterior *vs.* posterior) or another surgical procedure (*e.g.*, lumbar arthroplasty). The studies collectively included patients with an age of 18 to 71 years presenting with intractable back pain greater than leg pain and image-confirmed disc degeneration by such factors as disc space collapse, abnormal motion on flexion/extension radiographs, Modic endplate changes, and/or vertebral osteophyte formation. In 7 of the 12 studies, provocative discography was required in the diagnosis; others reported its use inconsistently.

Fusion techniques included 1- or 2-level instrumented and noninstrumented PLF, PLIF, transforaminal lumbar interbody fusion, ALIF, and combinations of these, as well as comparisons of open and less invasive approaches. Internal comparisons were made across techniques, approaches, and grafting materials. Minimum follow-up in each study was for 2 years, with isolated reports of longer-term outcomes.³¹

As a whole, these 12 studies including 1420 fusion patients showed a weighted average improvement in back pain of $36.5 \pm 17.2/100$ points (43.3% change; 95% CI, 31.5–55.1), in ODI function of $24.7 \pm 6.2/100$ (47.2% change; 95% CI, 41.3–53.1), and in SF-36 PCS of $12.7 \pm 3.1/100$ points (44.2% change, 95% CI, 37.1–51.4) (Figure 3). The average satisfaction rate was 67.0% (95% CI, 61.1–73.0), the fusion rate was 89.3% (95% CI, 84.4–94.2), and reoperation rate was 15.3% (95% CI, 10.7–19.8), reflective of Federal Drug Administration (FDA) Investigational Device Exemption (IDE) trial reporting of revisions, removals, supplemental fixation (index-level procedures, 9.7%), and reoperations (inclusive of adjacent-level procedures, 5.6%).^{26,28,31,32,35}

TABLE 3. Summary of Prospective Randomized Controlled Trials of Surgery Only (Level 1b)

| Article | Cohorts | n* | Lvls | Follow-up (mo) | Mean Age (yr) | Disco-gram Required | Baseline Back Pain (/100+) | Change in Back Pain (abs [%]) | Baseline ODI (/100) | Change in ODI (abs [%]) | Baseline SF-36 PCS (/100) | Change in SF-36 PCS (abs [%]) | Satisf/Redo Rate (%) | Fusion Rate (%) | Reop Rate (%) |
|--------------------------------|---------------------------------|-----|------|----------------|---------------|---------------------|----------------------------|-------------------------------|---------------------|-------------------------|---------------------------|-------------------------------|----------------------|-----------------|---------------|
| Boden et al ²⁷ | Inst. PLF (BMP2) | 11 | | | 57.6 | | NR | 25.5 (NR) | NR | NR | NR | NR | 60 | 100 | 18 |
| | PLF (BMP2) | 11 | 1 | 24 | 51.8 | N | NR | 52.0 (NR) | NR | 28.7 (NR) | NR | NR | 100 | 100 | 9.1 |
| | Inst. PLF (ICBG) | 5 | | | 52.9 | | NR | 31.0 (NR) | NR | NR | NR | NR | 60 | 40 | 0 |
| Burkus et al ²⁸ | Lap ALIF (BMP2) | 134 | | | 42.4 | | NR | NR | 52.3 | 33.6 (64) | 28.3 | 16.7 (59) | NR | 94.2 | 9.0 |
| | Open ALIF (BMP2) | 143 | | | 43.3 | N | NR | NR | 53.7 | 29.8 (55) | 27.7 | 14.7 (53) | NR | 94.5 | 12.6 |
| | Lap ALIF (ICBG) | 266 | 1 | 24 | 40.0 | | NR | NR | 46.5 | 23.8 (51) | 29.5 | 13.0 (44) | NR | 88.7 | 21.4 |
| | Open ALIF (ICBG) | 136 | | | 42.3 | | NR | NR | 55.1 | 31.3 (57) | 29.4 | 12.7 (43) | NR | 89.8 | 13.2 |
| Chung et al ²⁹ | Lap ALIF (CF cages, ICBG) | 22 | | | 49 | | 91 | 51.0 (56) | 41 | 16 (39) | NR | NR | 82 | 90.9 | NR |
| | Mini-open ALIF (CF cages, ICBG) | 22 | 1 | 24 | 50 | Y | 94 | 57.0 (61) | 43 | 20 (47) | NR | NR | 73 | 90.9 | NR |
| Haid et al ³² | PLIF (BMP2) | 34 | | | 46 | | NR | 90.0 (NR) | NR | 29.6 (NR) | NR | NR | 72.4 | 97.3 | 17.6 |
| | PLIF (ICBG) | 33 | | | 46 | | NR | 45.0 (NR) | NR | 24.9 (NR) | NR | NR | 80.0 | 77.8 | 18.2 |
| Sasso et al ³⁵ | ALIF (CTF cage + ICBG) | 77 | | | 41 | | NR | NR | 51.1 | 21.6 (42) | 28.30 | 11.5 (41) | NR | 97.0 | 27.3 |
| | ALIF (FRA + ICBG) | 62 | 1 | 24-48 | 41.2 | N | NR | NR | 52.7 | 21.2 (40) | 28.50 | 8.8 (31) | NR | 51.9 | 50.0 |
| Blumenthal et al ³⁶ | ALIF (CTF cage + ICBG) | 99 | | | 39.6 | | 71.8 | 34.3 (48) | 52.1 | 21.6 (41) | NR | NR | 53.1 | 90.9 | 9.1 |
| | TDR | 205 | 1 | 24 | 39.6 | Y | 72.0 | 40.8 (57) | 50.6 | 24.3 (48) | NR | NR | 73.7 | NA | 5.4 |
| McKenna et al ³⁴ | Inst ALIF (CTF cage + ICBG) | 41 | | | 41.4 | | 71 | 19.0 (27) | 54 | 6 (11) | 34 | 5 (15) | NR | NR | 9.8 |
| | Inst ALIF (FRA + ICBG) | 37 | 1-2 | 28 | 39 | Y | 72 | 12.0 (17) | 57 | 15 (26) | 27 | 17 (63) | NR | NR | 2.7 |

(Continued)

TABLE 3. (Continued)

| Article | Cohorts | n* | Lvs | Follow-up (mo) | Mean Age (yr) | Disco-gram Re-quired | Baseline Back Pain (/100+) | Change in Back Pain (abs [%]) | Baseline ODI (/100) | Change in ODI (abs [%]) | Baseline SF-36 PCS (/100) | Change in SF-36 PCS (abs [%]) | Satisfi/Redo Rate (%) | Fusion Rate (%) | Reop Rate (%) |
|--------------------------------|--------------------------------|------|---------|----------------|---------------|------------------------|----------------------------|-------------------------------|---------------------|-------------------------|---------------------------|-------------------------------|-----------------------|-----------------|---------------|
| Jenis et al ²³ | 360 (AGF) | 15 | 1-2 | 24 | 40.3 | Y | 75 | 27.0 (36) | 59.5 | 12.5 (21) | 27.7 | 6.4 (23) | NR | 89 | 6.7 |
| | 360 (ICBG) | 22 | | | 41.4 | | 77 | 29.0 (38) | 60.0 | 23.3 (39) | 29.0 | 9.1 (31) | NR | 85 | 0 |
| Zigler et al ²⁶ | 360 (ALIF FRA, Instr PLF ICBG) | 75 | 1 | 24 | 40.4 | Y | NR | 32 (NR) | 62.7 | 22.9 (37) | NR | NR | 69 | 97 | 5.4 |
| | TDR | 161 | | | 38.7 | | NR | 39 (NR) | 63.4 | 28.9 (46) | NR | NR | NR | 81 | NA |
| Berg et al ²⁵ | PLIF or PLF | 72 | 1-2 | 24 | 38.5 | N | 58.5 | 29.3 (50) | 41.2 | 18.2 (44) | NR | NR | 67 | 97 | 9.7 |
| | TDR | 80 | | | 40.2 | | 62.3 | 36.9 (59) | 41.8 | 21.8 (52) | NR | NR | NR | 71 | NA |
| Guyer et al ²¹ | ALIF (CTF cage + ICBG) | 43 | 1 | 60 | 38.8 | Y | NR | 40.0 (NR) | NR | 27.5 (NR) | NR | 12.3 (NR) | 72 | NR | 16.3 |
| | TDR | 90 | | | 40 | | NR | 38.7 (NR) | NR | 24.0 (NR) | NR | 12.6 (NR) | NR | 78 | NA |
| Delamarter et al ²⁰ | 360 (ALIF FRA, Instr PLF ICBG) | 60 | 2 | 24 | 41.8 | Y | 74.7 | 36.3 (49) | 64.8 | 26.1 (40) | 30.1 | 9.1 (30) | 62.1 | 77 | 8.3 |
| | TDR | 148 | | | 41.8 | | 75.7 | 43.8 (58) | 64.7 | 34.4 (53) | 29.5 | 14.4 (49) | NR | 78.2 | NA |
| Fusion | Total | 1420 | Average | Average | 43.9 | Weighted averaged age† | 72.5 | 36.5 (43.3) | 52.1 | 24.7 (47.2) | 29.0 | 12.7 (44.2) | 67.0 | 89.3 | 15.3 |
| | | | SD | SD | 5.3 | W-SD# | | 17.2 (18.0) | | 6.2 (12.1) | | 3.1 (12.1) | 10.5 | 10.9 | 10.4 |
| TDR | Total | 684 | Average | Average | 40.1 | Weighted averaged age† | 71.5 | 40.3 (57.5) | 56.4 | 27.2 (49.2) | 29.5 | 13.7 (48.8) | 76.6 | NA | 5.2 |
| | | | SD | SD | 1.1 | W-SD# | | 2.5 (1.2) | | 5.0 (3.5) | | 0.9 (NA) | 3.9 | NA | 2.7 |

*The number of patients is reflective of cohorts reported at last follow-up, not enrollment.

†VAS scores were converted to a 100-point score where necessary, for consistent interpretation.

#Where appropriate, outcomes measures were averaged weighted by individual study size. Weighted standard deviations (W-SD) across groups of studies are also presented for change in outcome measures over time.

FU indicates follow-up; ODI, Oswestry Disability Index; SF-36, Short-Form (36) Health Survey; PCS, physical component score; PLF, posterolateral fusion; IBF, interbody fusion; PLIF, posterior lumbar interbody fusion; ALIF, anterior lumbar interbody fusion; BMP, bone morphogenic protein; ICBG, iliac crest bone graft; CF, carbon fiber; CTF, cylindrical threaded fusion; TDR, total disc replacement; FRA, femoral ring allograft; NR, not reported; NA, not applicable; SD, standard deviation; W-SD, weighted standard deviation (weighted by study size); n, number of patients; lvs, levels; mo, months; yr, years; abs, absolute; Min, minimum; Max, maximum; Instr, instrumented; Lap, laparoscopic; abs, absolute value; Satisfi, satisfaction; Reop, reoperation (includes nonindex level procedures).

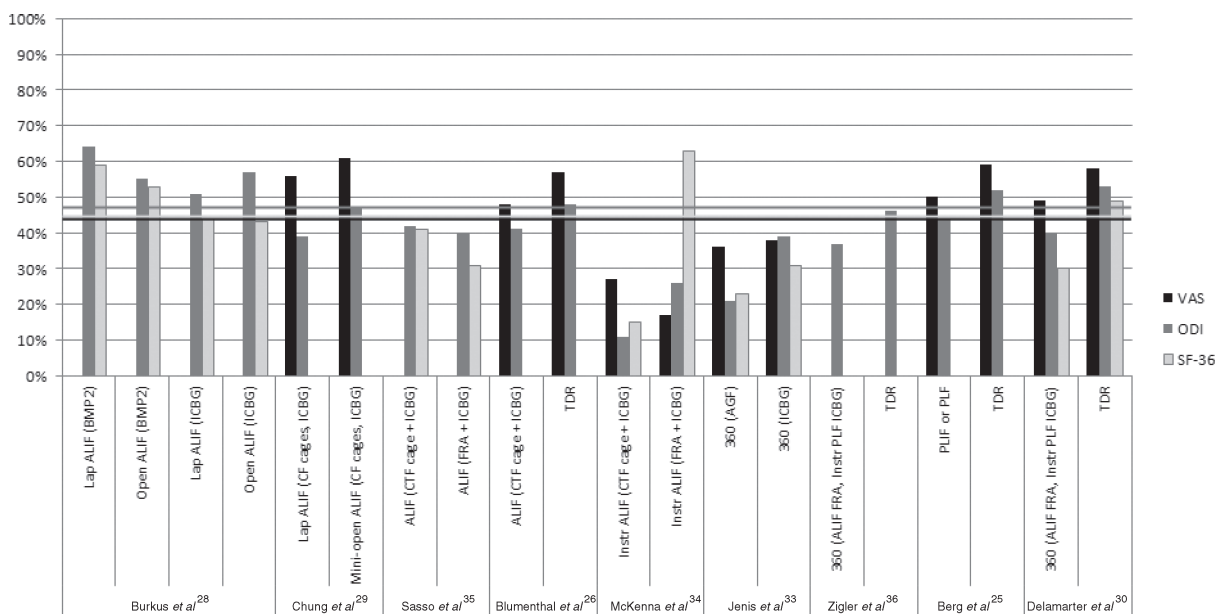


Figure 3. Graphical representation of reported percentage change (from preoperative to last follow-up) in back pain visual analogue scale (VAS), Oswestry Disability Index (ODI), and Short Form (36) Health Survey physical component score among randomized studies of surgery only (comparisons of procedures). ALIF indicates anterior lumbar interbody fusion; CF, carbon fiber; CTF, cylindrical threaded fusion; ICBG, iliac crest bone graft; BMP, bone morphogenic protein; FRA, femoral ring allograft; AGF, autologous growth factors; PLF, posterolateral fusion; PLIF, posterior lumbar interbody fusion; MIS, minimally invasive surgery.^{25–36}

Studies of Surgery-Only Cohorts: Prospective Nonrandomized Comparative Trials

Two publications^{37,38} (Table 4) report on the results of prospective trials with comparative cohorts, but which were not randomized. Both compare the results of open *versus* less invasive techniques and collectively included 381 patients, aged 24 to 74 years, undergoing 1- or 2-level lumbar fusion, with a minimum 2-year follow-up.

The weighted average improvement in back pain (VAS) was 48.0 ± 9.3/100 points (70.2% change; 95% CI, 59.2–81.2), in ODI was 26.9 ± 3.8/100 points (57.5% change; 95% CI, 49.9–65.1), and in SF-36 PCS was 17.2 ± 2.7/100 points (58.1% change; 95% CI, 46.1–70.2) (Figure 4). Patient satisfaction was not reported in these 2 studies. Weighted average fusion rate was 90.4% (95% CI, 86.2–94.5), and reoperation rate was not reported.

Studies of Surgery-Only Cohorts: Retrospective Comparative Trials

Two publications^{39,40} (Table 4) report on the results of retrospective studies with comparative cohorts. Fusion techniques included instrumented PLF, PLIF, transforaminal lumbar interbody fusion, ALIF, and circumferential 360° fusion. One study compared the results after fusion surgery at 1 to 2 levels *versus* 3 or more levels⁴⁰ but only the results of the 1 to 2 level cohort was included in the current review. Collectively, these 2 studies included 302 patients with a mean age of 46.8 years, undergoing 1- or 2-level lumbar fusion, with minimum 24 months of follow-up.

Back pain scores were not reported in these 2 studies. The weighted average improvement in ODI was 16.5 ±

11.8/100 points (32.3% change; 95% CI, 10.1–54.6), and in SF-36 PCS was 10.0 ± 2.8/100 points (41.9% change; 95%CI, not evaluable) (Figure 4). Patient satisfaction was not reported in these 2 studies. Fusion rate was 95.1% and reoperation rate was 2.5% in the single study reporting these results.

Studies of Surgery-Only Cohorts: Prospective Noncomparative Trials

Two publications^{41,42} (Table 4) report on the results of prospective single-cohort trials, that is, with no comparator. Fusion techniques included PLIF,⁴¹ and ALIF.⁴² Collectively, these 2 studies included results of 167 patients, aged 17 to 77 years, with minimum 1-year⁴¹ and 2-year⁴² follow-up.

The weighted average improvement in back pain (VAS) was 34.0/100 points (49.0% change; 95% CI, 26.4–71.5), in ODI was 19.0/100 points (49.1% change; 95% CI, 28.9–69.4), and in SF-36 PCS was 9.0/100 points (32% change; 95% CI not applicable, with only 1 article reporting) (Figure 4). Patient satisfaction averaged 77.9% (95% CI, 61.2–94.5). Fusion rate was 86.5% (95% CI, 61.1–100), and reoperation rate was 7.8% (95% CI, 0–21.4), mainly supplemental posterior fixation of stand-alone ALIF cages to promote fusion healing.⁴²

Studies of Surgery-Only Cohorts: Retrospective Noncomparative Trials

Two publications^{43,44} report on the results of retrospective noncomparative single-cohort studies. Fusion techniques included 1- or 2-level instrumented PLF and minimally

TABLE 4. Summary of Nonrandomized Surgery-Only Trials

| Article | Cohorts | N | Lvls | Follow-up (mo) | Mean Age (yr) | Discogram Required | Baseline Back Pain (/100+) | Change in Back Pain (abs [%]) | Baseline ODI (/100) | Change in ODI (abs [%]) | Baseline SF-36 PCS (/100) | Change in SF-36 PCS (abs [%]) | Satisf/Redo rate (%) | Fusion Rate (%) | Reop Rate (%) | |
|--|---------------------------|-----|---------|----------------|-------------------|--------------------|----------------------------|-------------------------------|---------------------|-------------------------|---------------------------|-------------------------------|----------------------|-----------------|---------------|----|
| Kasis et al ³⁷ | Open PLIF | 114 | 1-2 | 24 | 48.9 | N | 64 | 37 (58) | 46.4 | 22.5 (48) | 29.3 | 14.7 (50) | NR | 91 | NR | |
| | MIS PLIF | 209 | | | 45.6 | | 72 | 28.6 (61) | 29.6 | 18.5 (63) | NR | 92 | NR | | | |
| Peng et al ³⁸ | Open TLIF | 29 | 1-2 | 24 | 52.5 | N | 65 | 53 (82) | 47.7 | 30.2 (63) | NR | NR (NR) | NR | 86.7 | NR | |
| | MIS TLIF | 29 | | | 54.1 | | 60 | 29.0 (64) | NR | NR (NR) | NR | 80.0 | NR | | | |
| Total of prospective nonrandomized studies (level 2) | | 381 | Average | 50.3 | Weighted average† | 68.2 | 48.0 (70.2) | 46.8 | 26.9 (57.5) | 29.5 | 17.2 (58.1) | NR | 90.4 | NR | NR | |
| | | | SD | 3.8 | W-SD‡ | | 9.3 (11.2) | | 3.8 (7.8) | | 2.7 (8.7) | NA | 4.2 | NA | NA | |
| Lettec et al ⁴⁰ | PLIF and 360 (1-2 levels) | 81 | 1-2 | 24 | 41.0 | Y | NR | NR (NR) | NR | NR (NR) | 26.5 | 11.1 (42) | NR | 95 | 2.5 | |
| | | | | | | | | | | | | | | | | |
| Glassman et al ³⁹ | Overall (multitechniques) | 221 | | | 51.2 | | NR | NR (NR) | NR | 22.1 (NR) | NR | 9.5 (NR) | NR | NR | NR | |
| | | | | | 42 | | NR | NR (NR) | 47.8 | 27.0 (56) | NR | 13.8 (NR) | NR | NR | NR | |
| | | | | | 52 | | NR | NR (NR) | 55.9 | 25.1 (45) | NR | 11.2 (NR) | NR | NR | NR | NR |
| | | | | | 55 | | NR | NR (NR) | 46.1 | 4.0 (9) | NR | 7.7 (NR) | NR | NR | NR | NR |
| | 360 | 37 | | 44 | | W-SD‡ | NR | NR (NR) | 51.4 | 18.9 (37) | NR | 6.3 (NR) | NR | NR | | |
| Total of retrospective comparative studies (level 3) | | 302 | Average | 46.8 | Weighted average† | NR | NR (NR) | 49.8 | 16.5 (32.3) | 26.5 | 10.0 (41.9) | NR | 95.1 | 2.5 | | |
| | | | SD | 6.3 | W-SD‡ | NA | NA (NA) | | 11.8 (22.7) | | 2.8 (NA) | NA | NA | NA | NA | |
| Folman et al ⁴¹ | PLIF | 87 | NR | 12 | 45 | N | NR | NR (60) | 31.0 | 18.3 (59) | NR | NR (NR) | 86 | 98.9 | 1.1 | |
| | | | | | | | | | | | | | | | | |
| Li et al ⁴² | ALIF with I/F cage (ICBG) | 80 | 1 | 24 | 41.7 | N | 92 | 34 (37) | 51.3 | 19.7 (38) | 28.4 | 9.0 (32) | 69 | 73 | 15.0 | |
| | | | | | | | | | | | | | | | | |
| Total of prospective noncomparative studies (level 4a) | | 167 | Average | 43.4 | Weighted average† | 92.0 | 34.0 (49.0) | 40.7 | 19.0 (49.1) | 28.4 | 9.0 (32) | 77.9 | 86.5 | 8.1 | | |
| | | | SD | 2.3 | W-SD‡ | NA | NA (NA) | | 1.0 (14.6) | | NA (NA) | 12.0 | 18.3 | 9.8 | | |
| Schwender et al ⁴⁴ | MIS TLIF | 49 | 1 | 18 | NR | N | 72 | 51 (71) | 46 | 32 (70) | NR | NR (NR) | NR | 100 | 8.2 | |
| Dimar et al ⁴³ | Instr PLF (ICBG) | 194 | 1-2 | 24 | 52.3 | N | 79 | 40 (51) | 51.6 | 25.2 (49) | 27.4 | 12.3 (45) | NR | 89.3 | 13.9 | |
| | | | | | | | | | | | | | | | | |
| Total of retrospective noncomparative studies (level 4b) | | 243 | Average | 52.3 | Weighted average† | 77.6 | 42.2 (54.7) | 50.5 | 26.6 (53.0) | 27.4 | 12.3 (44.9) | NR | 91.5 | 12.7 | | |
| | | | SD | NA | W-SD‡ | | 7.8 (14.3) | | 4.8 (14.7) | | NA (NA) | NA | 7.6 | 4.1 | | |

(Continued)

TABLE 4. (Continued)

| Article | Cohorts | N | Lvlis | Follow-up (mo) | Mean Age (yr) | Discogram Required | Baseline Back Pain (/100+) | Change in Back Pain (abs [%]) | Baseline ODI (/100) | Change in ODI (abs [%]) | Baseline SF-36 PCS (/100) | Change in SF-36 PCS (abs [%]) | Satisf/Redo rate (%) | Fusion Rate (%) | Reop Rate (%) |
|-------------------------------------|---------|------|---------|----------------|---------------|--------------------|----------------------------|-------------------------------|---------------------|-------------------------|---------------------------|-------------------------------|----------------------|-----------------|---------------|
| Total across all levels of evidence | | 3060 | Average | | 46.3 | Weighted average† | 71.9 | 36.8 (50.4) | 49.5 | 22.2 (44.5) | 26.9 | 12.5 (46.5) | 71.1 | 89.1 | 12.5 |
| | | | SD | | 4.2 | W-SD# | | 14.7 (16.1) | | 14.1 (26.0) | | 4.3 (12.0) | 5.2 | 13.5 | 12.4 |

From top to bottom: prospective nonrandomized comparative studies (level 2), retrospective comparative studies (level 3), prospective noncomparative studies (level 4a), retrospective noncomparative studies (level 4b), and overall weighted averages across all levels of evidence inclusive of data in Tables 2 to 4.

*The number of patients is reflective of cohorts reported at last follow-up, not enrollment.

†VAS scores were converted to a 100-point score where necessary, for consistent interpretation.

#Where appropriate, outcomes measures were averaged weighted by individual study size. Weighted standard deviations (W-SD) across groups of studies are also presented for change in outcome measures over time.

N indicates number of patients; FU, follow-up; ODI, Oswestry Disability Index; PCS, physical component score; PLIF, posterior lumbar interbody fusion; MIS, minimally invasive surgery; PLF, posterolateral fusion; ICBG, iliac crest bone graft; TLIF, transforaminal lumbar interbody fusion; ALIF, anterior lumbar interbody fusion; BMP, bone morphogenic protein; I/F, interbody fusion; NR, not reported; NA, not applicable; SD, standard deviation; W-SD, weighted standard deviation (weighted by study size); SF-36, Short Form (36) Health Survey; Lvlis, levels; mo, months; yr, years; abs, absolute; Min, minimum; Max, maximum; abs, absolute value; Satisf, satisfaction; Reop, reoperation (includes nonindex level procedures); Lap, laparoscopic; Inst, instrumented.

invasive transforaminal lumbar interbody fusion, with results of 243 patients aged 18 to 86 years.

The weighted average improvement in back pain was $42.2 \pm 7.8/100$ points (54.7% change; 95% CI, 34.9–74.5), and in ODI was $26.6 \pm 4.8/100$ points (53.0% change; 95% CI, 32.7–73.3), and in SF-36 PCS was 12.3/100 points (44.9% change) in a single study⁴³ (Figure 4). Patient satisfaction was not reported in these 2 studies. Fusion rate was 91.5% (95% CI, 81.0–100), and reoperation rate was 12.7% (95% CI, 7.1–18.4), inclusive of removal of symptomatic and asymptomatic instrumentation and supplemental fixation of nonunions of posterior fusions.⁴³

Minimum Clinically Important Difference

Three publications^{45–47} sought to define the minimum amount of change in patient-reported outcomes scores that is clinically relevant, or MCID. For back pain scores, the MCID was defined as low as 12⁴⁶ and as high as 30⁴⁵ on a VAS scale of 100. For ODI, the suggested MCID has been reported between 12.8⁴⁶ and 22⁴⁵ points (on a 100-point scale). For SF-36 PCS scores, MCID has been suggested as an improvement of 4.9⁴⁶ to 6.2⁴⁵ points (on a 100-point scale).

The weighted averages from the combined 3060 lumbar fusion patients across the 26 studies evaluated met or exceeded MCID on all patient-reported variables. Back pain (VAS) improved by $36.8 \pm 14.7/100$ points (50.4% change; 95% CI, 43.6–57.1), ODI improved by $22.2 \pm 14.1/100$ points (44.5% change; 95% CI, 35.8–53.3), SF-36 PCS improved by $12.5 \pm 4.3/100$ points (46.5% change; 95% CI, 40.8–52.2). Additionally, patient satisfaction averaged 71.1% (95% CI, 68.6–73.7). Fusion rates averaged 89.1% (95% CI, 84.4–93.8). Overall reoperation rates averaged 12.5% (95% CI, 7.8–17.2), 9.2% of these at the index level, including supplemental fixation of stand-alone procedures and late elective hardware removal.

Analysis of variance showed that the average results of the collective studies do not differ statistically in any of the outcome measures based on the level of evidence of the study ($P > 0.05$). However, there was a statistical difference in the reoperation rates among procedure types, the highest being in stand-alone ALIF (19%), compared with posterior fusion without interbody support (10%), and the lowest being circumferential fusion (5%) ($P = 0.0082$). There were no statistical differences due to procedure with respect to patient-reported outcomes measures, satisfaction, or fusion rates.

DISCUSSION

Despite the publication of several systematic reviews on fusion surgery for DDD or chronic LBP,^{48–54} recommendations have been inconclusive. These reviews have generally relied on a limited number of prospective, randomized trials comparing fusion and nonsurgical treatments.^{50,52} In reality the comparison between fusion and nonsurgical treatment of LBP is artificial because in clinical practice, nonsurgical care and surgery are not competitive treatments. They are treatments usually performed in series rather than in parallel. Surgery is typically performed only after failure of

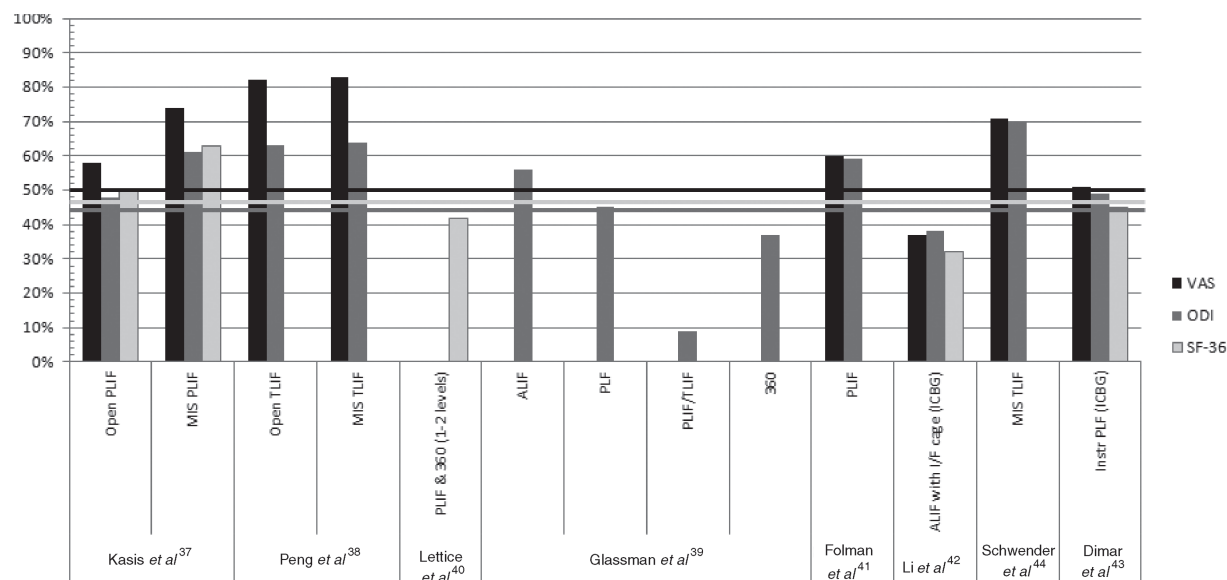


Figure 4. Graphical representation of reported percentage change (from preoperative to last follow-up) in back pain visual analogue scale (VAS), Oswestry Disability Index (ODI), and Short Form (36) Health Survey physical component score among nonrandomized studies of surgery only.³⁷⁻⁴⁴ Solid horizontal lines represent weighted averages across all cited studies. ICBG indicates iliac crest bone graft; PLF, posterolateral fusion; TLIF, transforaminal lumbar interbody fusion; MIS, minimally invasive surgery.

nonoperative treatment, limiting the value of comparisons between these treatments.

Review of the published level 1 studies is instructive in terms of highlighting the difficulties in relying exclusively on these for decision making. Fritzell *et al*²³ suggested greater improvement in back-specific disability after fusion compared with nonoperative care for LBP. On the other hand, 2 Norwegian studies by Brox *et al*^{19,20} suggested no substantial difference in disability when fusion was compared with intensive cognitive intervention and exercise rehabilitation. Mirza *et al*⁵² noted that these Norwegian studies were underpowered to identify clinically important differences. In addition, these studies had only 1 year of follow-up and also included postdiscectomy patients, and the conclusions may or may not apply to patients with LBP who have not had prior surgery. A British study²² of LBP treatment found that the pooled mean difference in ODI between the surgical and nonsurgical groups was in favor of surgery, but cautioned about the risk of complications with surgery.

Furthermore, exclusive reliance on only level 1 studies is confounding because blinded clinical trials of fusion surgery are inherently difficult to conduct because the patient and clinical investigator readily appreciate the intervention. The significant crossover in the Spine Patient Outcomes Research Trial exemplifies the difficulty of conducting an appropriate randomized surgical *versus* nonsurgical trial.⁵⁵⁻⁵⁸ As a significantly powered multicenter NIH-funded study, the authors have been applauded for the effort, although the results of the study have been challenged because of crossover rates from nonsurgical to surgical treatments, study design (the sequence and quality of nonoperative and operative care allocation),⁵⁹ and interpretation (intent-to-treat *vs.* as-treated analysis).^{60,61} Because of the challenges in study

design and execution, most surgical studies of LBP use the patients as their own control.

To better understand the scope of outcomes data available for clinical decision making, a broader appreciation of evidence-based medicine would include the collective body of research performed, understanding the limitations as study methodologies decrease in scientific rigor. Pioneer of evidence-based medicine David Sackett⁶² wrote, “the practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research... Evidence-based medicine is not restricted to randomized trials and meta-analyses. It involves tracking down the best external evidence with which to answer our clinical questions.” RCT design, through randomization, seeks to control and equilibrate unknown prognostic factors. However, generalization of RCT results to broader patient populations is difficult due to inclusionary restrictions and protocol parameters, which often do not reflect real-world medical practice.^{17,18} Thus, observational study design has been repeatedly described as complementary to RCTs in its ability to aid RCT results in making population conclusions by examining a broader, more inclusionary sample under normal practice conditions, especially when modeled to account for the inherent potential for bias.^{17,18,63-67}

Hartz *et al*⁶⁵ evaluated this issue and concluded that observational spinal fusion studies do complement spinal fusion RCTs and should be considered in decision making. This finding is consistent with the current review, where no statistical difference was found in any of the outcome measures between RCTs and observational studies based on the level of evidence ($P > 0.05$). This lends validity to the scope of the collective results and further support that the average improvements in

pain and function after fusion surgery for chronic LBP are reproducible across study designs.

The improvements in pain and function after surgery seem to be sustained during the long term. Burkus *et al*,⁶⁸ for example, published the 6-year results of the same patient cohort that was included in this systematic review with 2-year outcomes,²⁸ with equivalent or better outcomes compared with the 2-year results.⁶⁸ Furthermore, the clinical benefit of surgery does not seem to be dependent on the type of fusion procedure. In addition to the analysis of variance reported across the studies in the current review, an additional publication by Fritzell *et al*⁶⁹ on the cohort of patients included in their prospective, randomized trial,²³ found no significant differences between outcomes and the fusion techniques used (PLFs and interbody fusions). The current systematic review did however note that the lowest rate of reoperation was achieved with circumferential fusion (posterior instrumentation with interbody support). Stratifying the cohort reported by Dimar⁴³ by those older than or younger than 65 years, Glassman *et al*⁷⁰ reported that the benefits of lumbar fusion were similar or better in an older patient population, and concluded that treatment need not be withheld based on age.

One of the limitations of this or any literature review is the inconsistency of reporting methods and results of the studies reviewed. Consistent diagnostic criteria to determine the homogeneity of the patient populations within the reviewed studies was difficult to retrieve. However, studies included in this review specifically focused on chronic LBP as the primary complaint. All surgical patients had symptoms recalcitrant to nonoperative management, with documented radiographical confirmation of level-specific degeneration, without radiographical findings or clinical signs of neural compression, significant stenosis, deformity, fracture, infection, or neoplasm. Another variable among studies was the timing of surgery, although most studies required patients to have failed a minimum of some form of nonoperative therapy. Further studies are required to understand the optimal timing of surgery for these conditions and early identification of risk factors for persistence of pain in nonsurgically treated patients.

In terms of outcomes, validated patient-reported clinical outcomes measures were assessed. A limitation of the current review is that VAS pain scores were variably reported in the individual studies (some reported on a scale of 0–100 and others 0–10 or 0–20); however, all scores were converted to a 0 to 100 score to allow for comparisons across studies. Complication reporting was also challenging as individual studies defined complications differently, and reported varying degrees of adverse events. The more structured US FDA IDE studies included several complications, enumerating the incidences of each, but did not consistently report the percentage of patients in whom a complication occurred. Instead, reoperation rate was used as a more consistent measure of negative outcomes. However, categorization of secondary surgeries also varied by study, and included elective removal of instrumentation, planned secondary surgery for supplemental hardware, as well as adjacent level surgeries that might not be considered complications of the index surgery, but should

nevertheless be discussed with patients to advise that additional surgery (not necessarily revision surgery) may be necessary in up to 12% of cases.

Finally, MCID was used here as a reference point to compare average improvements within and across studies with reported thresholds for clinical significance. MCID is designed to report a percentage of individual patients who meet an improvement threshold. These data per patient were unavailable in the published studies, but MCID was instead used less formally to demonstrate a relative context for clinical improvements.

This systematic review analyzed currently published literature to evaluate the efficacy of fusion for chronic LBP with lumbar disc degeneration that was refractory to nonsurgical care. Improvement in pain and function were documented with the degree of clinical improvement comparable with that seen in other common, well-accepted orthopedic procedures such as total knee replacement, hip revision, and spinal decompression surgery for spinal stenosis.⁷¹ There is consistent evidence from these randomized and nonrandomized clinical studies that lumbar spine fusion decreases pain and disability in patients with chronic LBP related to degeneration of the motion segment.

➤ Key Points

- ❑ Review of the literature on fusion as a treatment of LBP has historically focused only on RCTs of surgery versus nonoperative treatments, which are limited.
- ❑ The full body of literature on the topic also includes RCTs between fusion and nonfusion surgical procedures and across fusion approaches, as well as prospective and retrospective nonrandomized studies that contribute additional real world findings.
- ❑ There is consistent evidence from these randomized and nonrandomized clinical studies reviewed, that lumbar spine fusion results in clinically meaningful improvements in pain and function, with acceptable patient satisfaction and low rates of revision in selected patients with chronic LBP related to degeneration of the motion segment.

References

1. Andersson GB. Epidemiological features of chronic low-back pain. *Lancet* 1999;354:581–5.
2. Deyo RA, Weinstein JN. Low back pain. *N Engl J Med* 2001;344:363–70.
3. Freburger JK, Holmes GM, Agans RP, et al. The rising prevalence of chronic low back pain. *Arch Intern Med* 2009;169:251–8.
4. Cassidy JD, Carroll LJ, Cote P. The Saskatchewan health and back pain survey. The prevalence of low back pain and related disability in Saskatchewan adults. *Spine* 1998;23:1860–6.
5. Elliott AM, Smith BH, Penny KI, et al. The epidemiology of chronic pain in the community. *Lancet* 1999;354:1248–52.
6. National Center for Health Statistics. [http://www.cdc.gov/nchs/data/05.pdf](http://www.cdc.gov/nchs/data/hus/05.pdf). Accessed November 5, 2011.
7. Katz JN. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. *J Bone Joint Surg Am* 2006;88(suppl 2):21–4.

8. Adams MA, Roughley PJ. What is intervertebral disc degeneration, and what causes it? *Spine* 2006;31:2151–61.
9. Edgar MA. The nerve supply of the lumbar intervertebral disc. *J Bone Joint Surg Br* 2007;89:1135–9.
10. Kallewaard JW, Terheggen MA, Groen GJ, et al. Discogenic low back pain. *Pain Pract* 2010;10:560–79.
11. Henchoz Y, Pinget C, Wasserfallen JB, et al. Cost-utility analysis of a three-month exercise programme vs usual care following multidisciplinary rehabilitation for chronic low back pain. *J Rehabil Med* 2010;42:846–52.
12. Wilson JF. In the clinic. Low back pain. *Ann Intern Med* 2008;148:ITC5.
13. Martell BA, O'Connor PG, Kerns RD, et al. Systematic review: opioid treatment for chronic back pain: prevalence, efficacy, and association with addiction. *Ann Intern Med* 2007;146:116–27.
14. Allen RT, Rihn JA, Glassman SD, et al. An evidence-based approach to spine surgery. *Am J Med Qual* 2009;24:15S–24S.
15. Freburger JK, Carey TS, Holmes GM. Physical therapy for chronic low back pain in North Carolina: overuse, underuse, or misuse? *Phys Ther* 2011;91:484–95.
16. Deyo RA, Gray DT, Kreuter W, et al. United States trends in lumbar fusion surgery for degenerative conditions. *Spine* 2005;30:1441–5.
17. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;342:1887–92.
18. Concato J. Observational versus experimental studies: what's the evidence for a hierarchy? *NeuroRx* 2004;1:341–7.
19. Brox JI, Sorensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine* 2003;28:1913–21.
20. Brox JI, Reikeras O, Nygaard O, et al. Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: a prospective randomized controlled study. *Pain* 2006;122:145–55.
21. Brox JI, Nygaard OP, Holm I, et al. Four-year follow-up of surgical versus non-surgical therapy for chronic low back pain. *Ann Rheum Dis* 2010;69:1643–8.
22. Fairbank J, Frost H, Wilson-MacDonald J, et al. Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *BMJ* 2005;330:1233.
23. Fritzell P, Hagg O, Wessberg P, et al. 2001 Volvo award winner in clinical studies: lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine* 2001;26:2521–32.
24. Ohtori S, Koshi T, Yamashita M, et al. Surgical versus nonsurgical treatment of selected patients with discogenic low back pain: a small-sized randomized trial. *Spine* 2011;36:347–54.
25. Berg S, Tullberg T, Branth B, et al. Total disc replacement compared to lumbar fusion: a randomised controlled trial with 2-year follow-up. *Eur Spine J* 2009;18:1512–9.
26. Blumenthal S, McAfee PC, Guyer RD, et al. A prospective, randomized, multicenter Food and Drug Administration investigational device exemptions study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: part I: evaluation of clinical outcomes. *Spine* 2005;30:1565–75.
27. Boden SD, Kang J, Sandhu H, et al. Use of recombinant human bone morphogenetic protein-2 to achieve posterolateral lumbar spine fusion in humans: a prospective, randomized clinical pilot trial: 2002 Volvo Award in clinical studies. *Spine* 2002;27:2662–73.
28. Burkus JK, Heim SE, Gornet MF, et al. Is INFUSE bone graft superior to autograft bone? An integrated analysis of clinical trials using the LT-CAGE lumbar tapered fusion device. *J Spinal Disord Tech* 2003;16:113–22.
29. Chung SK, Lee SH, Lim SR, et al. Comparative study of laparoscopic L5-S1 fusion versus open mini-ALIF, with a minimum 2-year follow-up. *Eur Spine J* 2003;12:613–7.
30. Delamarter R, Zigler JE, Balderston RA, et al. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement compared with circumferential arthrodesis for the treatment of two-level lumbar degenerative disc disease: results at twenty-four months. *J Bone Joint Surg Am* 2011;93:705–15.
31. Guyer RD, McAfee PC, Banco RJ, et al. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: five-year follow-up. *Spine J* 2009;9:374–86.
32. Haid RW Jr, Branch CL Jr, Alexander JT, et al. Posterior lumbar interbody fusion using recombinant human bone morphogenetic protein type 2 with cylindrical interbody cages. *Spine J* 2004;4:527–38.
33. Jenis LG, Banco RJ, Kwon B. A prospective study of Autologous Growth Factors (AGF) in lumbar interbody fusion. *Spine J* 2006;6:14–20.
34. McKenna PJ, Freeman BJ, Mulholland RC, et al. A prospective, randomised controlled trial of femoral ring allograft versus a titanium cage in circumferential lumbar spinal fusion with minimum 2-year clinical results. *Eur Spine J* 2005;14:727–37.
35. Sasso RC, Kitchel SH, Dawson EG. A prospective, randomized controlled clinical trial of anterior lumbar interbody fusion using a titanium cylindrical threaded fusion device. *Spine* 2004;29:113–22.
36. Zigler J, Delamarter R, Spivak JM, et al. Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine* 2007;32:1155–62.
37. Kasis AG, Marshman LA, Krishna M, et al. Significantly improved outcomes with a less invasive posterior lumbar interbody fusion incorporating total facetectomy. *Spine* 2009;34:572–7.
38. Peng CW, Yue WM, Poh SY, et al. Clinical and radiological outcomes of minimally invasive versus open transforaminal lumbar interbody fusion. *Spine* 2009;34:1385–9.
39. Glassman S, Gornet MF, Branch C, et al. MOS short form 36 and Oswestry Disability Index outcomes in lumbar fusion: a multicenter experience. *Spine J* 2006;6:21–6.
40. Lettice JJ, Kula TA, Derby R, et al. Does the number of levels affect lumbar fusion outcome? *Spine* 2005;30:675–81.
41. Folman Y, Lee SH, Silvera JR, et al. Posterior lumbar interbody fusion for degenerative disc disease using a minimally invasive B-twin expandable spinal spacer: a multicenter study. *J Spinal Disord Tech* 2003;16:455–60.
42. Li J, Dumonski ML, Liu Q, et al. A multicenter study to evaluate the safety and efficacy of a stand-alone anterior carbon I/F Cage for anterior lumbar interbody fusion: two-year results from a Food and Drug Administration investigational device exemption clinical trial. *Spine* 2010;35:E1564–70.
43. Dimar JR, Glassman SD, Burkus JK, et al. Two-year fusion and clinical outcomes in 224 patients treated with a single-level instrumented posterolateral fusion with iliac crest bone graft. *Spine J* 2009;9:880–5.
44. Schwender JD, Holly LT, Rouben DP, et al. Minimally invasive transforaminal lumbar interbody fusion (TLIF): technical feasibility and initial results. *J Spinal Disord Tech* 2005;18(suppl):S1–6.
45. Carragee EJ, Cheng I. Minimum acceptable outcomes after lumbar spinal fusion. *Spine J* 2010;10:313–20.
46. Copay AG, Glassman SD, Subach BR, et al. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. *Spine J* 2008;8:968–74.
47. Glassman SD, Copay AG, Berven SH, et al. Defining substantial clinical benefit following lumbar spine arthrodesis. *J Bone Joint Surg Am* 2008;90:1839–47.
48. Carreon LY, Glassman SD, Howard J. Fusion and nonsurgical treatment for symptomatic lumbar degenerative disease: a systematic review of Oswestry Disability Index and MOS Short Form-36 outcomes. *Spine J* 2008;8:747–55.
49. Chou R, Baisden J, Carragee EJ, et al. Surgery for low back pain: a review of the evidence for an American Pain Society Clinical Practice Guideline. *Spine* 2009;34:1094–109.

50. Gibson JN, Waddell G. Surgery for degenerative lumbar spondylosis: updated Cochrane Review. *Spine* 2005;30:2312–20.
51. Ibrahim T, Tleyjeh IM, Gabbar O. Surgical versus non-surgical treatment of chronic low back pain: a meta-analysis of randomised trials. *Int Orthop* 2008;32:107–13.
52. Mirza SK, Deyo RA. Systematic review of randomized trials comparing lumbar fusion surgery to nonoperative care for treatment of chronic back pain. *Spine* 2007;32:816–23.
53. Resnick DK, Choudhri TF, Dailey AT, et al. Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 7: intractable low-back pain without stenosis or spondylolisthesis. *J Neurosurg Spine* 2005;2:670–2.
54. Yajun W, Yue Z, Xiuxin H, et al. A meta-analysis of artificial total disc replacement versus fusion for lumbar degenerative disc disease. *Eur Spine J* 2010;19:1250–61.
55. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA* 2006;296:2441–50.
56. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT) observational cohort. *JAMA* 2006;296:2451–9.
57. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med* 2007;356:2257–70.
58. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med* 2008;358:794–810.
59. Spine Patient Outcome Research Trial (SPORT): multi-center randomized clinical trial of surgical and non-surgical approaches to the treatment of low back pain. *Spine J* 2003;3:417–9.
60. Angevine PD, McCormick PC. Inference and validity in the SPORT herniated lumbar disc randomized clinical trial. *Spine J* 2007;7:387–91.
61. Fisher C, Bishop P. SPORT and the Canadian experience. *Spine J* 2007;7:263–5.
62. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71–2.
63. *User's Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago, IL: AMA Press; 2002:218.
64. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;342:1878–86.
65. Hartz A, Benson K, Glaser J, et al. Assessing observational studies of spinal fusion and chemonucleolysis. *Spine* 2003;28:2268–75.
66. Hartz A, Bentler S, Charlton M, et al. Assessing observational studies of medical treatments. *Emerg Themes Epidemiol* 2005;2:8.
67. Ligthelm RJ, Borzi V, Gumprecht J, et al. Importance of observational studies in clinical practice. *Clin Ther* 2007;29:1284–92.
68. Burkus JK, Gornet MF, Schuler TC, et al. Six-year outcomes of anterior lumbar interbody arthrodesis with use of interbody fusion cages and recombinant human bone morphogenetic protein-2. *J Bone Joint Surg Am* 2009;91:1181–9.
69. Fritzell P, Hagg O, Wessberg P, et al. Chronic low back pain and fusion: a comparison of three surgical techniques: a prospective multicenter randomized study from the Swedish lumbar spine study group. *Spine* 2002;27:1131–41.
70. Glassman SD, Polly DW, Bono CM, et al. Outcome of lumbar arthrodesis in patients sixty-five years of age or older. *J Bone Joint Surg Am* 2009;91:783–90.
71. Hansson T, Hansson E, Malchau H. Utility of spine surgery: a comparison of common elective orthopaedic surgical procedures. *Spine* 2008;33:2819–30.

AMERICAN ASSOCIATION OF
NEUROLOGICAL SURGEONS

THOMAS A. MARSHALL, *Executive Director*
5550 Meadowbrook Drive
Rolling Meadows, IL 60008
Phone: 888-566-AANS
Fax: 847-378-0600
info@aans.org



President

PAUL C. MCCORMICK, MD, MPH
New York, New York



CONGRESS OF
NEUROLOGICAL SURGEONS

REGINA SHUPAK, *Acting Executive Director*
10 North Martingale Road, Suite 190
Schaumburg, IL 60173
Phone: 877-517-1CNS
FAX: 847-240-0804
info@1CNS.org

President

CHRISTOPHER E. WOLFLA, MD
Milwaukee, Wisconsin

February 21, 2012

Scientific Resource Center, Oregon EPC
Mail code: BICC
3181 S.W. Sam Jackson Park Road
Portland, Oregon 97239-3098

RE: Key Questions -- Spinal Fusion for Painful Lumbar Degenerative Disc or Joint Disease

To whom it concerns:

On behalf of the American Association of Neurological Surgeons (AANS) and the Congress of Neurological Surgeons (CNS), we would like to thank the Agency for Healthcare Research and Quality (AHRQ) for the opportunity to comment on the Key Questions regarding proposed research on the topic of ***“Spinal Fusion for Painful Lumbar Degenerative Disc or Joint Disease”***. We appreciate the efforts of AHRQ’s Effective Health Care Program, and the research summaries regarding the benefits and risks of different treatment options for health conditions based on comparative effectiveness reviews. We also understand that these research summaries are not clinical recommendations or guidelines, but are nevertheless frequently utilized as such with respect to healthcare policy development.

For the formulation of each of these Key Questions, AHRQ has requested a description of the included studies including patient indications, methods of diagnosis, inclusion and exclusion criteria, treatments, and surgical techniques and devices used. The AANS and CNS, along with other medical societies, have developed clinical guidelines on this topic and do not feel that another systematic review of these questions will yield useful information where our previous efforts have concluded that there is a paucity of sufficient data and that the quality of the studies is limited. However, as evidenced by the similar limitations in other medical and surgical topics, this does not diminish the benefit of this surgical treatment to our patients. Questions posed for the “Comment on Key Questions” may not be clinically relevant, which may be the genesis for the state of our current medical literature, and why future studies based on these Key Questions may not lead to improvements in patient care.

With these preliminary comments in mind, we will now turn our attention to commenting on the specific questions posed by AHRQ:

1. **For adults with low back pain attributed to degenerative disc disease of the lumbar spine, does spinal fusion differ from nonoperative treatment in the ability to improve:**
 - a. **Patient-centered outcomes such as function, quality of life, or pain?**
 - b. **Adverse events?**

AHRQ has proposed performing a systematic review of the comparative effectiveness and safety of lumbar fusion versus nonsurgical treatment for low back pain attributed to degenerative disc disease.

Currently, the primary treatment for most individuals with low back pain related to lumbar degenerative disease is non-operative therapy. As written, the question reflects a misunderstanding of the issue in that the population of patients treated with surgery is selected from those who have already failed extensive non-operative management. Viewing surgical and nonsurgical therapies as competing is inappropriate in this patient population as they are complementary, and surgery is typically not performed unless non-operative modalities have already failed. In this patient population, non-operative treatments have already been demonstrated to not improve outcomes.

In patients with chronic disabling pain refractory to conservative measures, lumbar fusion surgery is a potential therapeutic option. In this difficult patient population, prospective studies demonstrate a 36.0 - 63.9 percent reduction in back disability as measured by the Oswestry Disability Index (ODI) at 2 years after lumbar fusion (1, 2, 3, 4). Back pain scores also decrease 31.9 - 54.6 percent over the same duration (2, 3, 4). Further, lumbar fusion is associated with a 130.9 – 140.6 percent improvement in overall health as measured by the physical health component of the Medical Outcomes Study 36-item Short Form Health Survey (SF-36) (1).

To date, there are four multicenter randomized controlled trials comparing lumbar fusion surgery versus nonoperative treatment for low back pain attributed to degenerative disc disease. All four studies employed standardized patient-centered outcome measures to assess function and pain. The Swedish Lumbar Spine Study Group randomized patients who failed conservative therapy for ≥ 2 years to lumbar fusion surgery versus nonoperative therapy (ranging from physical therapy, education, transcutaneous electrical nerve stimulation, epidural steroid injections, cognitive and functional training, and/ or coping strategies) (5). Patients were evaluated for 2 years post treatment. The surgical group demonstrated a 33 percent reduction in back pain score and a 25 percent decrease in ODI. Sixty-three percent of surgical patients rated themselves as “much better” postoperatively, and 36 percent had returned to work. Comparatively, the nonsurgical group demonstrated only a 7 percent reduction in back pain score and a 6 percent decrease in ODI. Only 29 percent of nonsurgical patients rated themselves as “much better” after treatment, and only 13 percent had returned to work.

Brox et al randomized a much smaller group of patients with low back pain who had failed 1 year of conservative therapy to lumbar fusion versus a nonsurgical treatment protocol consisting of a lengthy inpatient program of physical therapy, cognitive intervention, education and peer counseling which is not available in North America (6). Patients were evaluated for 1 year post treatment. The surgical group demonstrated a 36.6 percent reduction in back pain score and a 37.1 percent decrease in ODI. Conversely, the nonoperative group demonstrated only a 24.0 percent reduction in back pain score and a 30.9 percent decrease in ODI. Overall, 71 percent of surgical patients rated their treatment as successful compared to 63 percent of nonoperative patients. In a similar study, Brox et al randomized patients with low back pain after prior disc herniation surgery to either of the same treatment arms (7). More modest improvements were observed overall with the lumbar fusion group demonstrating a 21.5 percent reduction in back pain score and an 18.9 percent decrease in ODI. The nonsurgical group demonstrated a 23.5 percent reduction in back pain and a 28.4% decrease in ODI.

Fairbank et al randomized patients with degenerative disc disease related low back pain to lumbar fusion surgery versus nonoperative therapy consisting of an intensive inpatient rehabilitation program of cognitive behavioral therapy and exercise (8). Patients were evaluated for 2 years post treatment. The study was plagued by a high rate of crossover and significant patient loss to follow-up which heavily biased the study against surgical intervention given the intent to treat study design. Another significant methodological flaw related to the surgical group. Many patients were treated without fusion, making any statements regarding the efficacy of fusion based on the data from this study highly suspect. Despite the inherent biases against surgical intervention, the surgical group demonstrated a 26.9 percent decrease in ODI compared to only a 19.4 percent decrease observed in the nonoperative group. Overall general

health was assessed via the physical component of the SF-36, with the surgical group demonstrating a 148.5 percent improvement compared to only a 138.0 percent increase seen in the nonoperative group. A recent paper reported the 6-year follow-up of an FDA Phase IV study, combining patients from sites of two previous FDA trials on anterior lumbar interbody fusion for patients with DDD unresponsive to conservative care. This study reported a substantial improvement in patient daily functioning, with improvements in back pain, leg pain, Oswestry disability index (ODI), and Short-form 36 (SF-36) measures (25).

Lumbar fusion surgery for low back pain however carries risk of potential adverse events. Depending on the series, incidences of major and minor complications widely vary. Complications including neurologic events, approach related vascular injuries, wound infection, deep venous thrombosis, pseudoarthrosis, dural tear, and bone graft donor site pain among others ranged from 7.9- 46.4 percent (1, 3, 5, 6, 7, 8). Reoperation rates also widely varied ranging from 7.8 - 37.4 percent (1, 2, 3, 5, 8). Mortality after lumbar fusion surgery in these series was 0 - 0.7 percent (1, 2, 3, 4, 5, 6, 7, 8).

The existing literature demonstrates that both nonsurgical treatment and lumbar fusion surgery may improve function and pain for individuals with low back pain attributed to degenerative disc disease. While limited evidence suggests that lumbar fusion may result in better outcomes compared to nonoperative treatment for certain individuals, several systematic reviews have debated these conclusions (9, 10, 11). In 2005, the American Association of Neurological Surgeons and the Congress of Neurological Surgeons performed a joint systematic review and concluded that there is Class I evidence to support lumbar fusion for carefully selected patients with low back pain intractable to the best medical management (12). They also found that Class III medical evidence suggests that nonsurgical treatment consisting of intensive cognitive and physical therapy may be an efficacious option for patients with chronic disabling low back pain. Given these current systematic reviews, it is unlikely that the AHRQ's proposed re-assessment of the present literature will provide any further clarification of the comparative effectiveness of surgical and nonsurgical treatment of low back pain attributed to lumbar degenerative disc disease.

- 2. For adults with low back pain attributed to degenerative (not congenital) stenosis of the lumbar spine, does spinal fusion differ from nonoperative treatment in the ability to improve:**
 - a. Patient-centered outcomes such as function, quality of life, or pain?**
 - b. Adverse events?**

Fusion is not recommended in patients operated upon for spinal stenosis in the absence of deformity (such as spondylolisthesis, scoliosis, or regional kyphosis) or instability (pre-existing or iatrogenic) (12). There is substantial evidence indicating that surgical intervention improves pain, function, and quality of life (44). There is further evidence that these improvements are durable and cost effective. The use of fusion in this population should be applied selectively to those patients with the above listed risk factors for progressive instability or deformity. There are no non-operative measures demonstrated to improve long term outcomes in patients with neurogenic claudication due to lumbar stenosis (57, 58).

The population of patients with low back covers rather extensive subgroups and diagnoses. As such, these patients are so heterogeneous that comparison of patient-centered outcomes (such as function, quality of life, adverse events, or pain) following spinal fusion versus non-operative management is an impractical task. Several primary and secondary confounding issues, such as return to work, disability requirements, perception bias of type of treatment and also long term and short term goals of the patient, clinical practitioner and medical payer, further cloud the evaluation of effectiveness of both treatment arms considered above (13, 14).

Over the last few decades, an awareness of the above variety of factors and patient demographics have resulted in recent multiple studies trying to elucidate the effect of the two treatment arms discussed with regard to sub populations of adults and also timing of intervention (15, 16, 17).

In designing questions related to patient outcomes, particularly in symptom and function dependent conditions such as lumbar stenosis, specific questions, pertaining to specific subgroup of patients beyond age (e.g. adult versus pediatric), gender, and diagnosis type (e.g. congenital versus degenerative) need to be clarified. It is impossible for current static low back pain classification systems geared toward short term outcomes accurately determine dynamic long term benefits (18, 19, 20).

With regard to guidelines and policies that are government-sponsored, patient-centered outcome studies and recommendations, there is heterogeneity of both medical specialty society recommendations and also that of the medical payer policies due to variations in the literature and also transparency in the development of the policies (21).

In formulating questions on patient-centered outcomes related to function, quality of life, pain or adverse events, due to the complexity of the subject, variation of beneficiaries and lack of effective long term data, it is important to have clearly identified subgroups and also quality studies across specialty/ society groups identifying specific outcomes to avoid erroneous generalizations.

- 3. For adults with low back pain attributed to degenerative spondylolisthesis of the lumbar spine, does spinal fusion differ from nonoperative treatment in the ability to improve:**
 - a. Patient-centered outcomes such as function, quality of life, or pain?**
 - b. Adverse events?**

Several studies have compared fusion surgery to non-operative treatment for the indication of degenerative spondylolisthesis. These studies have shown that for patients who suffer from low back pain due to degenerative spondylolisthesis, surgical intervention in the form of fusion surgery is more effective than non-operative treatment. Weinstein et al showed in the SPORT trial that surgical intervention for the treatment of degenerative spondylolisthesis showed significant improvement in SF-36 for bodily pain and physical function, as well as statistically significant improvement in the Oswestry Disability Index (29). These improvements were maintained for a follow-up of four years.

With regards to surgical complication rate, Sansur et al reviewed over 10,000 patients with degenerative and isthmic spondylolisthesis for complication incidence and factors associated with adverse events (28). The total rate of complications was 9.2 percent, and included dural tears, wound infections, hardware and implant complications, and neurological complications. Factors that correlated with a higher complication rate included higher grade spondylolisthesis, and age > 65 years old. Degenerative spondylolisthesis had a higher complication rate than isthmic spondylolisthesis (8.5 percent vs. 6.6 percent, $p=0.002$). These complication rates do not differ significantly from those in other series published in the literature (40, 41, 42, 43). The complication rate for patients undergoing surgical intervention for degenerative spondylolisthesis, while obviously higher than the complication rate of non-surgical treatment, are consistent with complication rates for spine surgery in general, and should not be a deterrent to pursuing surgical intervention, which provides longer term and more definitive treatment of back pain for degenerative spondylolisthesis.

Lumbar fusion has been shown in multiple studies in the literature to be a more effective treatment for degenerative spondylolisthesis, and provides improvement in pain and disability that is superior to conservative therapy.

- 4. For adults with low back pain attributed to degenerative disc disease of the lumbar spine, does spinal fusion differ from other spinal procedures (e.g., total disc replacement, disc decompression) in the ability to improve:**
- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?**
 - b. Patient-centered outcomes such as function, quality of life, or pain?**
 - c. Adverse events?**

It is unclear from well executed randomized prospective trials that there is any difference between lumbar arthroplasty and lumbar fusion in operative treatment of patients with lumbar degenerative disc disease (DDD). Approval of lumbar arthroplasty by the U.S. Food and Drug Administration was predicated upon establishing parity in clinical outcomes with the standard of care, lumbar fusion. The FDA used the criterion of non-inferiority as the foundation for approving lumbar arthroplasty devices for widespread use (25).

A prospective randomized comparative trial of lumbar arthroplasty versus lumbar fusion assigned 72 adult DDD patients to posterolateral fusion (PLF) or posterior lumbar interbody fusion (PLIF) at 1-2 levels. Back pain and ODI scores decreased significantly at 2-years. At 2-years, 76 percent of fusion patients were back to work part or full time and 67 percent were satisfied with their surgery (26). A meta-analysis performed by Bono and Lee reviewed all publications on non-revision fusion for lumbar DDD from during a 20 year period, encompassing over 2000 patients. They report good or excellent clinical outcomes were achieved in over 70 percent of those treated (27).

Disc decompression, dynamic stabilization, facet replacement and many other evolving technologies do not have substantial literature support to allow comment on the relative efficacy of these procedures compared to lumbar fusion.

There are significant complications which may occur in patients undergoing lumbar spine fusions. Previous reports have not found a significant difference between arthroplasty and arthrodesis study cohorts. Disc degeneration may occur in segments adjacent to fusions in the lumbar and cervical spine. It is unclear whether or not these areas of "juxtafusional" disease are caused by the neighboring fusion or if they represent the natural progression of the lumbar and cervical degenerative processes.

These well designed and well executed studies have not demonstrated any differences in patient outcomes. It seems unlikely that further investigations will be superior to these efforts. Observational patient registries may be one means to answer these questions.

- 5. For adults with low back pain attributed to degenerative stenosis of the lumbar spine, does spinal fusion differ from other spinal procedures (e.g., decompressive laminectomy and minimally invasive procedures, including those using devices) in the ability to improve:**
- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?**
 - b. Patient-centered outcomes such as function, quality of life, or pain?**
 - c. Adverse events?**

Degenerative stenosis has diverse etiologies, and for Key Question #5 we must assume that the question is restricted to patients without an underlying need for spinal fusion such as in patients with spinal deformity or spondylolisthesis. Low back pain associated with degenerative stenosis without spinal instability or expected iatrogenic instability, such as in patients with spinal deformity or spondylolisthesis, does not alter the recommendations of decompressive laminectomy alone with targeted use of medial facetectomies and foraminotomies, with or without discectomy. Decompressive laminectomy has been supported for superiority over non-operative therapy in degenerative stenosis by

studies such as the SPORT trial. This randomized, prospective trial indicated substantially greater improvement in pain and function through 4 years after decompressive surgery (44).

The 2005 AANS/CNS guidelines on this topic noted that spinal fusion procedures are associated with improved outcomes in patients with pre-operative evidence of spinal instability (45). Hopp and Tsou first introduced the impact of iatrogenic instability occurring during surgery due to extensive facetectomy necessary to achieve decompression in 1988 (46). Subsequent reports have supported the concept (47, 48). Fox et al reported extensive decompression at more than one level without concomitant arthrodesis was associated with worse outcomes following decompressive laminectomy for lumbar degenerative spinal stenosis (48). The AANS/CNS Guidelines for Lumbar Fusion formally endorsed spinal fusion in addition to decompressive laminectomy under those circumstances of iatrogenic instability (45).

Minimally invasive options for the treatment of lumbar degenerative stenosis have gained widespread use but its rapid evolution has made its evaluation a moving target. There is extensive literature on the clinical utility of minimally invasive surgery as a safe and effective for the treatment of degenerative lumbar stenosis. Studies have indicated that minimally invasive spine surgery and traditional open lumbar surgery have similar long-term patient outcomes in terms of pain and quality of life (52, 55, 56). Studies and meta-analyses on peri-operative factors have reported equivalence in complication rates for minimally invasive surgery, with minimally invasive surgery associated with a lower post-operative wound infection, less intra-operative blood loss, longer operative times, with overall no difference in long-term patient outcomes (50, 51, 52). Fourny et al reported a systematic review in 2010 indicating no difference in adverse events (rates of reoperation, dural tear, cerebrospinal fluid leak, nerve injury, and infection) between minimally invasive lumbar decompression and open surgery, with or without fusion (49). Two more recent literature review and cost analysis studies suggested lower infection rates (and lower associated costs) for minimally invasive surgery (53, 54).

Laminectomy and other decompressive procedures are not generally performed for the treatment of axial low back pain. These procedures are performed to treat claudication or radiculopathy, with lumbar fusions indicated if there is pre-operative or expected intra-operative iatrogenic spinal instability.

- 6. For adults with low back pain attributed to spondylolisthesis of the lumbar spine, does spinal fusion differ from other spinal procedures (e.g., repair, vertebrectomy) in the ability to improve:**
- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?**
 - b. Patient-centered outcomes such as function, quality of life, or pain?**
 - c. Adverse events?**

The main treatment options for adult spondylolisthesis are decompression with fusion. Treatment of spondylolisthesis with fusion is the most common approach, and is the most clearly documented surgical option in the literature. The largest series reported is from the Scoliosis Research Society, where they reported the results of 10,242 surgically treated cases of adult spondylolisthesis. Out of 10,242 patients, only 532 were treated without fusion (28). Complications rates in patients undergoing fusion versus those undergoing decompression alone were not significantly different (28). In the SPORT trial, the vast majority of patients in the surgical group (who had superior outcomes when compared to the non-operative group) had fusions (29). The reason why this disease is treated mostly through fusion is due to reported risks of deformity progression and chronic pain in patients treated without fusion. Herkowitz demonstrated a high failure rate after decompression without fusion, and better outcomes with fusion (30). Other studies also support fusion in the treatment of this disease over other surgical options (31, 32).

Direct repair of the fractured pars interarticularis (spondylolysis) without fusing adjacent segments is a potential treatment option, but is limited to very minimal degrees of slip in younger patients who would have a better chance for bone formation along the fractured pars. A few studies report direct repair of the fractured pars, but there are no well recognized studies comparing pars repair to fusion, as the circumstances under which one would actually be able to consider pars repair alone are rare (33, 34). As discussed in the question, vertebrectomy is mentioned as a possible surgical alternative. Vertebrectomy would be reserved for very rare and severe circumstances of spondylolisthesis from trauma or oncologic conditions. Again due to the relative rarity of such situations, it cannot even be considered as a comparable treatment option in the routine patient with back pain and or leg symptoms from spondylolisthesis.

Since fusion remains the dominant treatment of choice in this condition, and as it has repeatedly been shown that fusion has more optimal results than decompression alone, it may not be useful to check for differences in perioperative outcomes such as surgery time, blood loss, or length of hospital stay. More long term outcomes, such as re-operation rates and long term quality of life measures have demonstrated that fusion is the superior treatment. Other options such as direct repair of pars, and vertebrectomy are indicated in rare circumstances and hence are not to be considered as comparable entities.

- 7. For adults with low back pain attributed to degenerative disc disease of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation or graft material) differ in the ability to improve:**
- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?**
 - b. Patient-centered outcomes such as function, quality of life, or pain?**
 - c. Adverse events?**

Clinicians understand that more involved procedures, such as combined anterior/posterior fusions, generally entail longer surgery, greater blood loss, and longer hospital stays. They are usually employed, however, in selected patients who are thought, prospectively, to be at risk for a suboptimal outcome from an alternative procedure because of individual patient factors or particular aspects of the patient's pathology. Many of these important differences, such as osteoporosis, significant motion on flexion/extension radiographs, or segmental kyphosis, are not routinely identified and studied in directly comparative investigations. On the contrary, most RCTs and other studies strive to achieve or to demonstrate complete balance between treatment cohorts and therefore treat differences between patients as potential sources of bias rather than as possible key indicators of the likely benefit of one technique over another.

For example, in the treatment of spondylolisthesis, there are several fusion techniques commonly employed including non-instrumented fusion, posterior instrumentation with posterolateral fusion (PLF), posterior instrumentation with interbody fusion, or a combined anterior and posterior approach. Each of these approaches has a role in the treatment of a heterogeneous patient population. An elderly patient with a collapsed disc space and a relatively fixed deformity would likely do well with a non-instrumented fusion whereas a younger patient with a more mobile spine would be at high risk for failure of that fusion construct and would be better treated with a more aggressive approach. The influence of spinal alignment, local anatomical features, osteoporosis, and patient demand (i.e. activity level and age) cannot be overstated. Evidence to this point is provided by Soegaard et al who found that circumferential fusion (the most costly and morbid) was associated with significant benefits and cost savings compared to less aggressive techniques in a working population (Soegaard et al: Circumferential fusion is dominant over posterolateral fusion in a long term perspective. Spine 32: 2405-2411, 2007). It is quite possible, indeed likely, that this benefit would not be apparent in an older patient population.

- 8. For adults with low back pain attributed to degenerative stenosis of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation or graft material) differ in the ability to improve:**
- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?**
 - b. Patient-centered outcomes such as function, quality of life, or pain?**
 - c. Adverse events?**

The response for Key Question #8 mirrors the discussion of Key Question #2. There are diverse indications for fusion in the setting of stenosis, and the approach varies with the diverse pathology and involved patient population. The use of fusion in the setting of stenosis is typically considered when instability is demonstrated pre-operatively or anticipated based on preoperative/intraoperative factors. In these circumstances, lumbar fusion has been shown to be beneficial, with improved function, quality of life, and pain. For symptomatic spinal stenosis with or without degenerative spondylolisthesis, a recent systematic review by Chou et al. found evidence that decompressive surgery is moderately superior to nonsurgical therapy through 1 to 2 years. Surgery for radiculopathy in the setting of symptomatic spinal stenosis is associated with short-term benefits compared to nonsurgical therapy, though benefits diminish with long-term follow-up in some trials. For nonradicular back pain with common degenerative changes, fusion is no more effective than intensive rehabilitation, but is associated with small to moderate benefits compared to standard nonsurgical therapy (10).

As highlighted in other Key Question responses, spinal fusions of any nature can increase surgery time, blood loss, potential for adverse events, and length of hospital stay, in comparison with simple decompression. It is understood by physicians that combined anterior-posterior fusion surgery will typically result in greater intraoperative time, blood loss, and length of hospital stay, and higher risk for adverse events – and that it is typically reserved for patients felt to be at risk for poor outcomes via a more limited approach (so as to improve functional or quality outcomes than would otherwise be expected). The superiority of a particular approach (anterior, posterior, combined) or technique (instrumentation or graft material) has not been proven, as the factors involved in a surgeon's decision are heterogeneous; options for approach are not always equal/competitive. Surgical techniques and approaches are constantly being refined. A study trying to prove superiority of one approach is doomed to limited relevance and will undoubtedly be an immense undertaking with likely equivocal outcomes.

- 9. For adults with low back pain attributed to spondylolisthesis of the lumbar spine, do spinal fusion approaches (e.g., anterior, posterior, combined) and techniques (e.g., instrumentation or graft material) differ in the ability to improve:**
- a. Perioperative outcomes such as surgery time, blood loss, or length of hospital stay?**
 - b. Patient-centered outcomes such as function, quality of life, or pain?**
 - c. Adverse events?**

The question put forth by AHRQ regarding the relative efficacy of the various spinal fusion approaches to address low back pain in patients with spondylolisthesis is far too broad a question in the expansive diagnosis of spondylolisthesis to conclusively answer. While examination of the various surgical approaches for a single diagnosis may seem at first glance appear to be a valid question for a homogeneous cohort, in reality spondylolisthesis is far from uniform. This diagnosis has within it various subsets and anatomical considerations that make it a heterogeneous group and therefore difficult to study.

For example, in the treatment of spondylolisthesis, several fusion techniques are commonly employed including non-instrumented fusion, posterior instrumentation with posterolateral fusion (PLF), posterior

instrumentation with interbody fusion, or a combined anterior and posterior approach. Each of these approaches has a role in the treatment of a heterogeneous patient population. An elderly patient with a collapsed disc space and a relatively fixed deformity would likely do well with a non-instrumented fusion whereas a younger patient with a more mobile spine would be at high risk for failure of that fusion construct and would be better treated with a more aggressive approach. The influence of spinal alignment, local anatomical features, osteoporosis, and patient demand (i.e. activity level and age) cannot be overstated. Evidence to this point is provided by Soegaard et al who found that circumferential fusion (the most costly and morbid) was associated with significant benefits and cost savings compared to less aggressive techniques in a working population (Soegaard et al: Circumferential fusion is dominant over posterolateral fusion in a long term perspective. *Spine* 32: 2405-2411, 2007). It is quite possible, indeed likely, that this benefit would not be apparent in an older patient population.

The largest and most expensive trial to date is the NIH funded Spine Patient Outcomes Research Trial (SPORT). While this trial represents the most comprehensive study to date examining the 3 common fusion methods used in the treatment of degenerative spondylolisthesis, it was not specifically designed to evaluate the three fusion techniques of posterolateral in situ fusion, posterolateral fusion with pedicle screw fixation and 360° fusion (PLIF/TLIF, ALIF augmented with pedicle screw stabilization). Regardless, this trial represents the largest cohort of degenerative spondylolisthesis available for review. The preliminary SPORT data demonstrated that individuals with spinal stenosis and associated degenerative spondylolisthesis treated surgically had substantially greater improvement in pain and function during a period of 4 years than did patients treated nonoperatively (29, 35). A subsequent evaluation of fusion methods within the same study attempted to examine the outcomes of 3 different fusion techniques: PLF, PPS and 360° fusion, but were unable to establish superiority of one approach over another. This is not because the procedures are equivalent, it is because they were each applied in appropriate patient populations and were generally successful.

With regards to the perioperative outcomes of surgery time and blood loss, times ranged from 157 to 274 minutes, with PLF having the shortest operative time and 360° having the longest. Mean blood loss ranged from 499 to 666 ml, again with PLF averaging the lowest and PPS averaging the highest. The most common adverse event was a dural tear, which was highest for PPS (12%) followed by PLF (9%) and lowest in 360° (2%). Incidentally, the rate of an inadvertent durotomy in this report seemed inordinately high. By comparison, Williams and colleagues reported a durotomy rate of 1.9 percent in patients with spondylolisthesis in their review of 108,478 cases (36). The postoperative transfusion rate in the SPORT study followed the same trend, PPS (26%), 360° (17%) and PLF (14%).

With regards to patient centered outcomes, all three groups' demonstrated significant improvement compared to baseline in various validated outcome measures (ODI, SF-36 BP and BF). There was no significant difference between the groups at 4 years (37). It is again important to emphasize that the SPORT study was not specifically designed to evaluate fusion techniques or to validate one form of fusion for the management of degenerative spondylolisthesis. While prospective in design, there was no randomization and therefore the results may have been affected by selection bias. Only a prospective randomized study designed and appropriately powered to evaluate these three fusion techniques in a narrow population with specific anatomical criteria has the capacity to determine which fusion method provides the greatest improvement in outcome measures and is the most cost effective treatment. However, the SPORT data has demonstrated the effectiveness of surgical treatment compared with nonsurgical treatment of degenerative spondylolisthesis.

While there is a constellation of reports in the literature that explore some element of the various subsets of question 9, there is no comprehensive study that unequivocally answers this question and, for the various reasons listed above, we do not foresee such a study ever taking place. What the literature has

unequivocally demonstrated is that surgeons have effectively used all three of these approaches to successfully treat patients with spondylolisthesis.

10. Are there patient characteristics (e.g., pain severity, prior treatment) that are associated with better or worse outcomes after spinal fusion?

- a. Patient-centered outcomes such as function, quality of life, or pain
- b. Adverse events

Some patient characteristics may have an effect on outcomes after spinal arthrodesis for lumbar degenerative disease. However, to date, no study has determined definitive preoperative characteristics which may predict optimal or suboptimal outcomes from lumbar arthrodesis. Several smaller studies and meta-analyses have reported preoperative parameters which may be included in the overall evaluation when considering a patient as a candidate for lumbar arthrodesis.

For example, psychiatric comorbidities have been examined as a potential predictor of outcomes. A recent meta-analysis evaluated outcomes from both nonsurgical and fusion treatments to examine the effect of psychiatric comorbidities on outcomes. While there were few studies specifically addressing this question, those studies suggested that patient whose comorbidities include a personality disorder, depression, or neuroticism should preferentially be treated non-operatively (15). Others have corroborated that the presence of depression may be an independent predictor of success for surgery (20). However, as Daubs et al. report, the strength of their recommendation is weak. While there are no definitive studies that would preclude surgery as an option for patients with psychiatric comorbidities, the studies cited suggest that it should be evaluated during decision making.

Other factors have been looked at as well, including preoperative health status, cardiac comorbidity, and work status among others. Preoperative health status self-assessment appears to be the most robust, yet definitive criteria for predicting outcome have not been established (38). Other factors such as radiographic findings have been explored as well. In general, when findings such as spondylolisthesis are present, these have been reported to portend a better outcome (9).

Overall, current literature does not support criteria or strong recommendations for excluding spinal arthrodesis due to specific preoperative patient characteristics (39).

Conclusion

We appreciate the opportunity to comment on the Key Question formulation regarding the AHRQ proposed research on the topic of “***Spinal Fusion for Painful Lumbar Degenerative Disc or Joint Disease***”. The AANS and CNS developed clinical guidelines on this topic in 2005, and we are currently undergoing the process of updating these guidelines. Based on our experience, we do not believe that another systematic review of these questions will yield useful information as there is a paucity of sufficient data and the quality of the studies is limited. **After reviewing the current literature in conjunction with the clinical expertise of our Neurosurgeon members, the AANS and CNS do not believe that this diminishes the benefit of this surgical treatment to our patients.** While we understand that these AHRQ research summaries are not clinical recommendations or guidelines, we remained concerned that this research proposal will involve a large effort with minimal and limited clinical relevance.


Again, thank you for this opportunity to comment and we look forward to seeing your final position pertaining to this proposed research. If you have any questions, please feel free to contact Joseph

Cheng, MD (joseph.cheng@vanderbilt.edu), AANS/CNS Committee for Payor and Policy Responses, or Koryn Rubin, the AANS/CNS Senior Manager for Quality Improvement.

Sincerely,



Paul C. McCormick, MD, MPH, President
American Association of Neurological Surgeons



Christopher E. Wolfla, MD, President
Congress of Neurological Surgeons

Staff Contact:

Koryn Rubin
Senior Manager, Quality Improvement
American Association of Neurological Surgeons/
Congress of Neurological Surgeons
Washington Office
725 15th Street, NW, Suite 500
Washington, DC 20005
Phone: 202-446-2030
Fax: 202-628-5264
E-mail: krubin@neurosurgery.org

References

1. Sasso, R.C., S.H. Kitchel, and E.G. Dawson, A prospective, randomized controlled clinical trial of anterior lumbar interbody fusion using a titanium cylindrical threaded fusion device. *Spine (Phila Pa 1976)*, 2004. 29(2): p. 113-22; discussion 121-2.
2. Burkus, J.K., et al., Clinical and radiographic outcomes of anterior lumbar interbody fusion using recombinant human bone morphogenetic protein-2. *Spine (Phila Pa 1976)*, 2002. 27(21): p. 2396-408.
3. Blumenthal, S., et al., A prospective, randomized, multicenter Food and Drug Administration investigational device exemptions study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: part I: evaluation of clinical outcomes. *Spine (Phila Pa 1976)*, 2005. 30(14): p. 1565-75; discussion E387-91.
4. Zigler, J., et al., Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. *Spine (Phila Pa 1976)*, 2007. 32(11): p. 1155-62; discussion 1163.
5. Fritzell, P., et al., 2001 Volvo Award Winner in Clinical Studies: Lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine (Phila Pa 1976)*, 2001. 26(23): p. 2521-32; discussion 2532-4.
6. Brox, J.I., et al., Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine (Phila Pa 1976)*, 2003. 28(17): p. 1913-21.
7. Brox, J.I., et al., Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: a prospective randomized controlled study. *Pain*, 2006. 122(1-2): p. 145-55.

8. Fairbank, J., et al., Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *BMJ*, 2005. 330(7502): p. 1233.
9. Carreon, L.Y., S.D. Glassman, and J. Howard, Fusion and nonsurgical treatment for symptomatic lumbar degenerative disease: a systematic review of Oswestry Disability Index and MOS Short Form-36 outcomes. *Spine J*, 2008. 8(5): p. 747-55.
10. Chou, R., et al., Surgery for low back pain: a review of the evidence for an American Pain Society Clinical Practice Guideline. *Spine (Phila Pa 1976)*, 2009. 34(10): p. 1094-109.
11. Mirza, S.K. and R.A. Deyo, Systematic review of randomized trials comparing lumbar fusion surgery to nonoperative care for treatment of chronic back pain. *Spine (Phila Pa 1976)*, 2007. 32(7): p. 816-23.
12. Resnick, D.K., et al., Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 7: intractable low-back pain without stenosis or spondylolisthesis. *J Neurosurg Spine*, 2005. 2(6): p. 670-2.
13. Radcliff KE, Rihn J, Hilibrand A, et al. Does the duration of symptoms in patients with spinal stenosis and degenerative spondylolisthesis affect outcomes?: Analysis of the spine outcomes research trial. *Spine (Phila Pa 1976)*. 2011;36(25):2197-2210. doi: 10.1097/BRS.0b013e3182341edf.
14. Choma TJ, Schuster JM, Norvell DC, Dettori JR, Chutkan NB. Fusion versus nonoperative management for chronic low back pain: Do comorbid diseases or general health factors affect outcome? *Spine (Phila Pa 1976)*. 2011;36(21 Suppl):S87-95. doi: 10.1097/BRS.0b013e31822ef89e.
15. Daubs MD, Norvell DC, McGuire R, et al. Fusion versus nonoperative care for chronic low back pain: Do psychological factors affect outcomes? *Spine (Phila Pa 1976)*. 2011;36(21 Suppl):S96-109. doi: 10.1097/BRS.0b013e31822ef6b9.
16. Mroz TE, Norvell DC, Ecker E, Gruenberg M, Dailey A, Brodke DS. Fusion versus nonoperative management for chronic low back pain: Do sociodemographic factors affect outcome? *Spine (Phila Pa 1976)*. 2011;36(21 Suppl):S75-86. doi: 10.1097/BRS.0b013e31822ef68c.
17. Boakye M, Moore R, Kong M, Skirboll SL, Arrigo RT. Health-related quality-of-life status in veterans with spinal disorders. *Qual Life Res*. 2012. doi: 10.1007/s11136-012-0121-y.
18. Fairbank J, Gwilym SE, France JC, et al. The role of classification of chronic low back pain. *Spine (Phila Pa 1976)*. 2011;36(21 Suppl):S19-42. doi: 10.1097/BRS.0b013e31822ef72c.
19. Sekiguchi M, Wakita T, Otani K, et al. Development and validation of a symptom scale for lumbar spinal stenosis. *Spine (Phila Pa 1976)*. 2012;37(3):232-239. doi: 10.1097/BRS.0b013e318216afb4.
20. Adogwa O, Parker SL, Shau DN, et al. Preoperative zung depression scale predicts outcome after revision lumbar surgery for adjacent segment disease, recurrent stenosis, and pseudarthrosis. *Spine J*. 2011. doi: 10.1016/j.spinee.2011.08.014.
21. Cheng JS, Lee MJ, Massicotte E, et al. Clinical guidelines and payer policies on fusion for the treatment of chronic low back pain. *Spine (Phila Pa 1976)*. 2011;36(21 Suppl):S144-63. doi: 10.1097/BRS.0b013e31822ef5b4.
22. Weinstein, J., et al., Surgical versus nonoperative treatment for lumbar disc herniation: Four-year results for the Spine Patient Outcomes Research Trial (SPORT). *Spine*, 2008. 33: p. 2789-2800.
23. Weinstein, J., et al., Surgical vs. nonoperative treatment for lumbar disk herniation: The Spine Patient Outcomes Research Trial (SPORT): A randomized trial. *JAMA*, 2006. 296: p. 2441-2450.
24. Hagg, O. and P. Fritzell, Re: Brox JI, Sörensen R, Friis A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine*. 2003; 28:1913-1921. *Spine*, 2004. 29: p. 1160-1161.
25. Geisler, F., et al., Neurological complications of lumbar artificial disc replacement and comparison of clinical results with those related to lumbar arthrodesis in the literature: results of a multicenter,

- prospective, randomized investigational device exemption study of Charité intervertebral disc. Invited submission from the Joint Section Meeting on Disorders of the Spine and Peripheral Nerves, March 2004. *Journal of Neurosurgery: Spine*, 2004. 1: p. 143-154.
26. Berg, S., et al., Total disc replacement compared to lumbar fusion: a randomised controlled trial with 2-year follow-up. *Eur Spine J*, 2009. 18: p. 1512-1519.
 27. Bono, C. and C. Lee, Critical Analysis of Trends in Fusion for Degenerative Disc Disease Over the Past 20 Years: Influence of Technique on Fusion Rate and Clinical Outcome. *Spine*, 2004. 29: p. 455-463.
 28. Sansur CA, Reames DL, Smith JS, Hamilton DK, Berven SH, Broadstone PA, Choma TJ, Goytan MJ, Noordeen HH, Knapp DR, Hart RA, Zeller RD, Donaldson WF, Polly DW, Perra JH, Boachie-Adjei O, Shaffrey CI: Morbidity and mortality in the surgical treatment of 10,242 adults with spondylolisthesis. *J Neurosurg Spine* 13:589-593.
 29. Weinstein JN, Lurie JD, Tosteson TD, Zhao W, Blood EA, Tosteson AN, Birkmeyer N, Herkowitz H, Longley M, Lenke L, Emery S, Hu SS: Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *J Bone Joint Surg Am* 91:1295-1304, 2009.
 30. Herkowitz HN, Kurz LT: Degenerative lumbar spondylolisthesis with spinal stenosis. A prospective study comparing decompression with decompression and intertransverse process arthrodesis. *J Bone Joint Surg Am* 73:802-808, 1991.
 31. Martin BI, Mirza SK, Comstock BA, Gray DT, Kreuter W, Deyo RA: Reoperation rates following lumbar spine surgery and the influence of spinal fusion procedures. *Spine (Phila Pa 1976)* 32:382-387, 2007.
 32. Martin CR, Gruszczynski AT, Braunsfurth HA, Fallatah SM, O'Neil J, Wai EK: The surgical management of degenerative lumbar spondylolisthesis: a systematic review. *Spine (Phila Pa 1976)* 32:1791-1798, 2007.
 33. Giudici F, Minoia L, Archetti M, Corriero AS, Zagra A: Long-term results of the direct repair of spondylolisthesis. *Eur Spine J* 20 Suppl 1:S115-120.
 34. Mutchnick IS, Clegg TE, Carreon LY, Puno RM: Motion segment-sparing repair of symptomatic chronic pars defects. *J Neurosurg Spine* 15:159-163.
 35. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *The New England journal of medicine*. May 31 2007;356(22):2257-2270.
 36. Williams BJ, Sansur CA, Smith JS, et al. Incidence of unintended durotomy in spine surgery based on 108,478 cases. *Neurosurgery*. Jan 2011;68(1):117-123; discussion 123-114.
 37. Abdu WA, Lurie JD, Spratt KF, et al. Degenerative spondylolisthesis: does fusion method influence outcome? Four-year results of the spine patient outcomes research trial. *Spine*. Oct 1 2009;34(21):2351-2360.
 38. Katz JN, Stucki G, Lipson MD, Et al. Predictors of Surgical Outcome in Degenerative Lumbar Spinal Stenosis. *Spine* 1999 24 2229-2233.
 39. Gibson JN, Waddell G Surgery for Degenerative Lumbar Spondylosis: Updated Cochrane Review. *Spine* 2005 20 2312-2320.
 40. DiPaolo CP, Molinari RW: Posterior lumbar interbody fusion. *J Am Acad Orthop Surg* 16:130-139, 2008.
 41. Ekman P, Mooler H, Tullbert T, Neuman P, Hedlund R: Posterior Lumbar interbody fusion versus posterolateral fusion in adult isthmic spondylolisthesis. *Spine* 32: 2178-2183, 2007.
 42. Fischgrund JS, Mackay M, Herkowitz HN, Brower R, Montgomer DM, Kurz LT: 1997 Volvo Award winner in clinic studies. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal instrumentation. *Spine* 22:2807-2812, 1997.

43. Glassman SD, Carreon LY, Djurasovic M, Dimar JR, Johnson JR, Puno RM et al: Lumbar fusion outcomes stratified by specific diagnostic indication. *Spine J* 9:13-21, 2009.
44. Weinstein JN et al Surgical versus nonoperative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial. *Spine (Phila Pa 1976)*. 35(14):1329-38, 2010.
45. Resnick DK et al, Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 10: fusion following decompression in patients with stenosis without spondylolisthesis. *J Neurosurg Spine*2(6):686-91, 2005.
46. Hopp E, Tsou P: Postdecompression lumbar instability. *Clin Orthop Relat Res* 227:143–151, 1988.
47. Brunon J, Chazal J, Chirossel J, et al: When is spinal fusion warranted in degenerative lumbar spinal stenosis? *Rev Rhum Engl Ed* 63:44–50, 1996.
48. Fox MW, Onofrio BM, Onofrio BM, et al: Clinical outcomes and radiological instability following decompressive lumbar laminectomy for degenerative spinal stenosis: a comparison of patients undergoing concomitant arthrodesis versus decompression alone. *J Neurosurg* 85:793–802, 1996.
49. Fourney et al. Does minimal access tubular assisted spine surgery increase or decrease complications in spinal decompression or fusion? *Spine (Phila Pa 1976)*. 35(9 Suppl):S57-65, 2010.
50. Park Y, Ha JW. Comparison of one-level posterior lumbar interbody fusion performed with a minimally invasive approach or a traditional open approach. *Spine (Phila Pa 1976)*. 32(5):537-43, 2007.
51. Ntoukas V, Müller A. Minimally invasive approach versus traditional open approach for one level posterior lumbar interbody fusion. *Minim Invasive Neurosurg*. 53(1):21-4, 2010.
52. Karikari IO, Isaacs RE. Minimally invasive transforaminal lumbar interbody fusion: a review of techniques and outcomes. *Spine (Phila Pa 1976)*. 35(26 Suppl):S294-301, 2010.
53. Parker SL et al. Post-operative infection after minimally invasive versus open transforaminal lumbar interbody fusion (TLIF): literature review and cost analysis. *Minim Invasive Neurosurg*. 54(1):33-7, 2011.
54. McGirt MJ et al. Comparative analysis of perioperative surgical site infection after minimallyinvasive versus open posterior/transforaminal lumbar interbody fusion: analysis of hospital billing and discharge data from 5170 patients. *J Neurosurg Spine*. 14(6):771-8, 2011.
55. Schizas C et al. Minimally invasive versus open transforaminal lumbar interbody fusion: evaluating initial experience. *Int Orthop*. 33(6):1683-8, 2009.
56. Wang J et al. Minimally invasive or open transforaminal lumbar interbody fusion as revision surgery for patients previously treated by open discectomy and decompression of the lumbar spine. *Eur Spine J*. 20(4):623-8, 2011.
57. Resnick DK, Choudhri TF, Dailey AT, Groff MW, Khoo L, Matz PG, Mummaneni P, Watters WC, Wang J, Walters BC, Hadley MN: Guidelines for the performance of lumbar fusion for degenerative disease of the lumbar spine. Part 13: Injection therapies for degenerative lumbar spine disease. *J Neurosurg Spine*2(6): 705-713, 2005
58. Watters WC III, Baisden J, Gilbert T, Kreiner DS, Resnick DK, Bono C, Ghiselli G, Heggeness M, Mazanec D, O'Neill C, Reitman C, Shaffer W, Summers J, Toton J: Diagnosis and treatment of degenerative lumbar stenosis. *Spine J* 8(2): 305-310, 2008



March 17, 2014

Health Care Authority
626 8th Ave SE
P.O. Box 42712
Olympia, WA 98504-2712

Attention: Christine Masters at shtap@hca.wa.gov

Dear Members of the Washington State Health Care Authority:

On behalf of the American Society for Metabolic and Bariatric Surgery (ASMBS), I am pleased to provide comments regarding the February 28, 2014 notice regarding proposed topics for review by the Washington State Health Care Authority's (HCA) Health Technology Assessment (HTA) program. Our comments will focus on bariatric surgery for those affected by obesity.

As you know, many public and private health plans provide coverage for bariatric surgery. For example, on the national level, the Medicare program began covering bariatric surgery in early 2006 based on its general coverage criteria that these services are "reasonable and necessary" for Medicare beneficiaries. Bariatric surgery is also covered by TRICARE and the Federal Employees Health Benefits Plan. On the state level, nearly every State Medicaid program (except for Montana and Ohio) provide coverage for bariatric surgery and over 75 percent of all active state public employees, such as those in Washington State, have access to coverage for bariatric surgery.

For these reasons, we are somewhat concerned that the HCA has chosen to review the safety, efficacy and costs associated with bariatric surgery – especially given the mounting evidence, which speak favorably to each of the aforementioned criteria. As this evidence continues to grow regarding the benefits of metabolic surgery –even in lower body mass index (BMI) populations (between 30 and 34.9 kg/m²), payers are starting to take notice.

For example, both the Hawaii Medical Service Association and the Cleveland Clinic Employee Health Plan provide coverage for metabolic surgery (Roux-en-Y gastric bypass) for those with a BMI of between 30 and 34.9 kg/m² with type II diabetes. In addition, an October, 2012 assessment by the Blue Cross Blue Shield Association's Technology Evaluation Center entitled, "Bariatric Surgery in Patients with Diabetes

and Body Mass Index Less than 35 kg/m²” found sufficient evidence supporting the use of gastric bypass for the treatment of type 2 diabetes in this population.

Following are our specific comments regarding the safety, efficacy and cost-savings associated with bariatric surgery:

SAFETY

Data involving nearly 60,000 bariatric patients from ASMBS Bariatric Centers of Excellence database show that the risk of death within the 30 days following bariatric surgery averages 0.13 percent, or approximately one out of 1,000 patients. This rate is considerably less than most other operations, including gallbladder and hip replacement surgery. Therefore, in spite of the poor health status of bariatric patients prior to surgery, the chance of dying from the operation is exceptionally low.

In addition, large studies find that the risk of death from any cause is considerably less for bariatric patients throughout time than for individuals affected by severe obesity who have never had the surgery. In fact, the data show up to an 89 percent reduction in mortality, as well as highly significant decreases in mortality rates due to specific diseases. Cancer mortality, for instance, is reduced by 60 percent for bariatric patients. Death in association with diabetes is reduced by more than 90 percent and mortality from heart disease by more than 50 percent. Also, there are numerous studies that have found improvement or resolution of life-threatening obesity-related diseases following bariatric surgery. The benefits of bariatric surgery, with regard to mortality, far outweigh the risks.

EFFICACY

Bariatric surgery, such as gastric bypass, gastric sleeve, and laparoscopic adjustable gastric banding, work by changing the anatomy of the gastrointestinal tract (stomach and digestive system) or by causing different physiologic changes in the body that change an individual’s energy balance and fat metabolism.

Severe obesity is one of the most serious stages of obesity. More than two decades ago, the National Institutes of Health (NIH) reported that individuals affected by severe obesity are resistant to maintaining weight loss achieved by conventional therapies, such as consuming fewer calories, increasing exercise, commercial weight-loss programs, etc.). The NIH recognized bariatric surgery as the only effective treatment to combat severe obesity and maintain weight loss in the long term.

Bariatric surgery is associated with massive weight-loss and improves, or even resolves, obesity-related co-morbidities for the majority of patients. These co-morbidities include high blood pressure, sleep apnea, asthma and other obesity-

related breathing disorders, arthritis, lipid (cholesterol) abnormalities, gastroesophageal reflux disease, fatty liver disease, venous stasis, urinary stress incontinence, pseudotumor cerebri, and more.

Bariatric surgery also leads to improvement and remission of Type II diabetes mellitus (T2DM). In the past, diabetes was considered to be a progressive and incurable disease. Treatments include weight loss and lifestyle changes for those who are overweight or obese and anti-diabetic medication, including insulin. These treatments help to control T2DM but rarely cause remission of the disease.

However, there is now a large body of scientific evidence showing remission of T2DM following bariatric surgery. A large review of 621 studies involving 135,247 patients found that bariatric surgery causes improvement of diabetes in more than 85 percent of the diabetic population and remission of the disease in 78 percent. In addition to improvements in health and longevity, surgical weight-loss improves overall quality of life. Measures of quality of life that are positively affected by bariatric surgery include physical functions such as mobility, self-esteem, work, social interactions, and sexual function. Singlehood is significantly reduced, as is unemployment and disability. Furthermore, depression and anxiety are significantly reduced following bariatric surgery.

COST SAVINGS ASSOCIATED WITH BARIATRIC SURGERY

From a cost-benefit analysis perspective, with the dramatic decline in co-morbid conditions, bariatric surgery has an average return on investment at approximately 24 months for laparoscopic procedures. It has also been determined that post-operative medication expenses for hypertension and diabetes medication drop by 77.3 percent.

In other health outcomes research studies, it has been determined that employer sponsored health insurers recoup their costs for bariatric surgery in two years when the procedure is performed as a laparoscopic surgery. This analysis not only included the surgery cost, but also the pre-surgical evaluation process, and five years post-bariatric surgery. This study was impactful because actual health insurance medical claims data were utilized – providing for a more reliable determination of real costs.

EXAMING METABOLIC SURGERY IMPACT ON TYPE 2 DIABETES

Finally, the ASMBS would like to offer the following Clinical Practice Guidelines and consensus statements that speak favorably about metabolic surgery's profound impact on patients with type 2 diabetes and a BMI < 35 kg/m²:

Clinical Practice Guidelines

- American Association of Clinical Endocrinologists, the Obesity Society, American Society for Metabolic and Bariatric Surgery (ASMBS): *Clinical Practice Guidelines for the Perioperative Nutritional, Metabolic, and Nonsurgical Support of the Bariatric Surgery Patient– 2013 Update*. 2013. This guideline states “patients with BMI of 30-34.9 kg/m² with diabetes or metabolic syndrome may also be offered a bariatric procedure, although current evidence is limited by the number of subjects studied and lack of long-term data demonstrating net benefit.”
- International Diabetes Federation: *Bariatric Surgical and Procedural Interventions in the Treatment of Obese Patients with T2DM*. 2011. This position statement states that “surgery should be considered as an alternative treatment option in patients with a BMI between 30 and 35 when diabetes cannot be adequately controlled by optimal medical regimen, especially in the presence of other major cardiovascular disease risk factors.”
- ADA: *Standards of Medical Care in Diabetes*. 2011. In this position statement, ADA recommends that “although small trials have shown glycemic benefit of bariatric surgery in patients with T2DM and BMI 30-35, there is currently insufficient evidence to generally recommend surgery in patients with BMI <35 outside of a research protocol.”
- American Heart Association: *Bariatric Surgery and Cardiovascular Risk Factors: A Scientific Statement From the American Heart Association*. 2011. In this document, the authors examine metabolic surgery for patients with BMI 30 to 35 kg/m² because of the poor results of nonoperative weight loss regimens, and they suggest that additional long-term data are needed before surgery for this group of patients becomes standard practice. The statement concludes, “at the moment, bariatric surgery should be reserved for patients who have severe obesity in whom efforts at medical therapy have failed and an acceptable operative risk is present.”
- International Federation for the Surgery of Obesity (IFSO), Asian Pacific Chapter (APC): *IFSO-APC Consensus Statements 2011*. This statement indicates that metabolic surgery “should be considered for the treatment of T2DM or metabolic syndrome for patients who are inadequately controlled by lifestyle alterations and medical treatment for acceptable Asian candidates with BMI ≥30.” It also indicates that metabolic surgery “may be considered as a non-primary alternative to treat inadequately controlled T2DM, or metabolic syndrome, for suitable Asian candidates with BMI ≥27.5.”
- Diabetes Surgery Summit (DSS): *The DSS Consensus Position Statement Recommendations for the Evaluation and Use of Gastrointestinal Surgery to Treat T2DM*. 2010. The position statement concludes that “there was strong

consensus that adopting the strictly BMI based criteria for metabolic surgery would be inadequate to select candidates for diabetes surgery.” The document further states that DSS recognizes value for metabolic surgery in “carefully selected, moderately obese patients (BMI 30-35) who are inadequately controlled by conventional medical and behavioral therapies.”

- Society of American Gastrointestinal and Endoscopic Surgeons: *Guidelines for Clinical Application of Laparoscopic Bariatric Surgery*. 2008. This document provides guidance for surgical treatment of obesity, patient selection, surgical procedure selection, programs, and preoperative preparation. The document also provides specific guidance for the different laparoscopic surgical procedures. The guidance document states that “individuals with BMI 30-35 kg/m² may benefit from laparoscopic bariatric surgery.”

In summary, ASMBS strongly supports continued coverage of bariatric surgery for Washington State employees and is hopeful that the HCA will examine the growing evidence surrounding the benefits of metabolic surgery for those with a BMI between 30 and 34.9 kg/m².

Should you have any questions or need additional information, please feel free to contact either me, or ASMBS Washington Policy Consultant Christopher Gallagher at 571-235-6475 or via email at chris@potomaccurrents.com. Thank you.

Sincerely,



Ninh Nguyen, MD, FACS, FASMBS
President, American Society for Metabolic and Bariatric Surgery



Brian Sung, MD, FASMBS
Swedish Weight Loss Services
Seattle, Washington



Judy Chen, MD, FASMBS
Swedish Weight Loss Services
Seattle, Washington